EÖTVÖS LORÁND UNIVERSITY
FACULTY OF EDUCATION AND PSYCHOLOGY

Eszter Tóth-Fáber

LEARNING AND CONSOLIDATION OF SKILLS IN TYPICAL AND ATYPICAL DEVELOPMENT

Doctoral School of Psychology
Head of the School: **Róbert Urbán,** PhD, DSc, professor, Eötvös Loránd University

Clinical Psychology and Addiction Program
Head of the Program: **Zsolt Demetrovics,** PhD, DSc, professor, Eötvös Loránd University

Supervisors
Dezső Németh, PhD, DSc, professor, Eötvös Loránd University
Karolina Janacsek, PhD, assistant professor, Eötvös Loránd University

Budapest, 2023

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

ADATLAP a doktori értekezés nyilvánosságra hozatalához és a DOI-azonosítók feltüntetéséről

I. A doktori értekezés adatai

A szerző neve: Tóth-Fáber Eszter

A doktori értekezés címe és alcíme: Learning and consolidation of skills in typical and atypical development

A doktori iskola neve: Pszichológiai Doktori Iskola

A doktori iskolán belüli doktori program neve: Klinikai pszichológia és addiktológia program

A témavezető neve és tudományos fokozata: Németh Dezső, PhD, DSc és Janacsek Karolina, PhD

A témavezető munkahelye: Eötvös Loránd Tudományegyetem

MTA Adatbázis-azonosító: 10068054

DOI-azonosító[1]: 10.15476/ELTE.2023.082

II. Nyilatkozatok

1. A doktori értekezés szerzőjeként[2]

a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom az ELTE PPK Doktori Iskola hivatalának ügyintézőjét Dávid Gergőt, hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltse a feltöltéshez szükséges nyilatkozatokat.

b) kérem, hogy a mellékelt kérelemben részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;[3]

[1] A kari hivatal ügyintézője tölti ki.

[2] A megfelelő szöveg aláhúzandó.

[3] A doktori értekezés benyújtásával egyidejűleg be kell adni a tudományági doktori tanácshoz a szabadalmi, illetőleg oltalmi bejelentést tanúsító okiratot és a nyilvánosságra hozatal elhalasztása iránti kérelmet.

c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (…….*dátum*)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;[4]
d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéséig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adatait tegyék közzé. Ha a könyv a fokozatszerzést követőn egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.[5]
2. A doktori értekezés szerzőjeként kijelentem, hogy
a) a ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti, önálló szellemi munkám és legjobb tudomásom szerint nem sértem vele senki szerzői jogait;
b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.
3. A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.

Kelt: Budapest, 2023.05.05.

a doktori értekezés szerzőjének aláírása

[4] A doktori értekezés benyújtásával egyidejűleg be kell nyújtani a minősített adatra vonatkozó közokiratot.
[5] A doktori értekezés benyújtásával egyidejűleg be kell nyújtani a mű kiadásáról szóló kiadói szerződést.

Table of contents

Acknowledgements

It is a truth universally acknowledged that one does not write a PhD dissertation by themselves. My dissertation is not an exception, and it could not have happened without many people. This page is dedicated to them, as they deserve more gratitude than I can ever express.

My supervisors, *Dezső Németh* and *Karolina Janacsek* welcomed me with open arms when I joined their lab. I am extremely grateful for their guidance and for sharing their passion for research with me. I greatly appreciate all the inspiring conversations, the shared social programs, conferences, lunches, and coffee breaks with *Dezső*. He always soothed my anxiety when it was needed and our conversations often helped me see academic life and my research in a brighter light. *Karolina* always inspired me to be a better researcher and was a great role model for her theoretical and methodological knowledge. Her thorough comments always raised the level of my publications and my work became better because of her. I consider myself lucky to have them as my mentors, and it gave me great comfort that I could turn to them with any problem or question. They supported me through all these years, and I could not have finished this dissertation without them.

My honorary supervisor, *Ádám Takács* showed me the beauty of research during my bachelor studies, helped me at the start of my academic career, and continued to offer me guidance throughout my PhD studies. I did not imagine that taking his class on ADHD would lead to this but I am grateful that it did. I was lucky enough to work with *Andrea Kóbor* as well, whom I always considered a role model for her persistence, perseverance, and work ethic. I would like to thank her for all the inspiring conversations regarding research and life in general and for the guidance she showed me. I appreciate all the joint work with *Zsanett Tárnok* whose expertise raised the level of the dissertation.

The past and present members of the Brain, Memory and Language Lab made the working days inspiring and amazing, and I am thankful that we maneuvered through the difficulties of the doctoral program together. *Zsófia Zavecz* was the best office mate I could ask for, even though we distracted each other just as often as we helped each other. I am glad we evolved from coworkers to friends. I am grateful to *Kata Horváth* for her friendship and for all the support she offered me from the start. I greatly appreciate all the help from *Orsolya Pesthy* as well, especially during the last few months of finishing the dissertation.

I was lucky to mentor and work with several great students during my PhD. I appreciate all their hard work. I would like to highlight *Bence Farkas, Flóra Hann, Dániel Sörnyei, Krisztina Berta* and *Bianka Brezóczki.* I am happy I can call all of them my friends now.

I received tremendous support from outside my lab as well. *István Tóth-Király* and *Beáta Bőthe* supported me at the start of my academic career, and I am thankful for all the meaningful conversations. I am also grateful to *Mónika Koós, Gábor Orosz* and *János Salamon* for showing me how much fun research can be. Lunchtime will never be the same without you.

Sharing an office space with the Budapest Laboratory of Sleep and Cognition resulted in several exciting conversations. I especially appreciate the regular check-ins and conversations with *Richárd Reichardt*, which always made me smile.

I would also like to thank my family, especially *Ákos Tóth-Fáber* and *Orsolya Horváth* and the people closest to me for offering never-ending support and patience during all these years and for listening to me ramble about my research projects at any given moment. *Borbála Tölgyesi* closely followed my research career from the first day of university, and the constant encouragement she offered me helped me through all the low points. *Bettina Wittman* arrived one week earlier for my birthday party in the 4th grade, and little did we know that would result in almost two decades of friendship, still counting. Her kindness lifted me up on bad days. My friendship with *Petra Major* started at the beginning of my PhD studies, and I know it will last long after that. She offered me endless help across countries and even continents. *János Hegedűs* got a crash course in the last eight months in the highs and lows of academic life and did excellently in supporting me during this challenging period. My loved ones were always there to celebrate even my smallest victories and I could not have done this work without their constant support.

Thank you all!

List of publications included in the dissertation

Tóth-Fáber, E., Tárnok, Z., Janacsek, K., Kóbor, A., Nagy, P., Farkas, B. C., Oláh, Sz., Merkl, D., Hegedűs, O., Nemeth, D., & Takács, Á. (2021). Dissociation between two aspects of procedural learning in Tourette syndrome: Enhanced statistical and impaired sequence learning. *Child Neuropsychology*, 27(6) 799-821. https://doi.org/10.1080/09297049.2021.1894110

Tóth-Fáber, E., Tárnok, Z., Takács, Á., Janacsek, K., & Nemeth, D. (2021). Access to procedural memories after one year: evidence for robust memory consolidation in Tourette syndrome. *Frontiers in Human Neuroscience*, 458. https://doi.org/10.3389/fnhum.2021.715254

Tóth-Fáber, E., Janacsek, K. & Nemeth, D. (2021). Statistical and sequence learning lead to persistent memory in children after a one-year offline period. *Scientific Reports, 11,* 12418. https://doi.org/10.1038/s41598-021-90560-5

Tóth-Fáber, E., Nemeth, D. & Janacsek, K. (2023). Lifespan developmental invariance in memory consolidation: evidence from procedural memory. *PNAS Nexus*, pgad037. https://doi.org/10.1093/pnasnexus/pgad037

Total impact factor of the published studies: 10.79

Note: Each co-author has granted permission for the given publication to be included in the current dissertation.

List of abbreviations

ADHD Attention Deficit Hyperactivity Disorder

ANOVA Analysis of Variance

ASRT Alternating Serial Reaction Time

BF Bayes Factors

CBGTC Cortico-Basal Ganglia-Thalamo-Cortical

CBIT Comprehensive Behavioral Intervention for Tics

EEG Electroencephalography

EF Executive Functions

EHI Edinburgh Handedness Inventory

LQ Laterality Quotient

LSD Least Significant Difference

M Mean

OCD Obsessive-Compulsive Disorder

RSE Residual Standard Error

RT Reaction Time

SD Standard Deviation

SDQ Strengths and Difficulties Questionnaire

SEM Standard Error of the Mean

SES Socioeconomic Status

SRT Serial Reaction Time

TD Typically Developing

TS Tourette Syndrome

WCST Wisconsin Card Sorting Test

YGTSS Yale Global Tic Severity Scale

I. General introduction

Acquiring and executing skills occurs regularly in everyday life. During the human lifespan, we learn several skills, especially at a young age. How we learn, consolidate, and retrieve such skills is highly relevant not only in typical development, but in atypical development as well. In my dissertation, I will focus on these questions and aim to provide a deeper understanding of skill learning, more precisely, of procedural memory, by investigating how the functional changes in fronto-striatal circuits in typical and atypical development might influence the learning and consolidation of procedural information.

In the general introduction, at first, I will present the multiple memory system with a focus on procedural memory and describe in detail the different regularities within procedural memory. This will be followed by a methodological section describing different versions of the Alternating Serial Reaction Time (ASRT) task, which were used in the studies of the dissertation to measure procedural memory. In the second half of the general introduction, I will summarize past research on procedural memory in typically and atypically developing children and across the lifespan. At last, the research questions of the respective studies included in the dissertation will be presented.

1. Multiple memory systems and phases of memory

Memory is not a unified construct; considering long-term memory, multiple memory systems can be distinguished (Squire, 1994; Squire & Wixted, 2011). The traditional taxonomy differentiates between declarative and non-declarative memory systems (however, for an alternative approach, see Henke, 2010). Declarative memory underlies learning and remembering facts (semantic memory) and events (episodic memory), and the memory representations are consciously accessible, flexible and affects behavior in different contexts. Taken the neural background, declarative memory is mainly related to the medial temporal lobe, especially to the hippocampus (Squire & Wixted, 2011). Non-declarative memory consists of several different memory types, such as priming, classical conditioning, emotional and perceptual learning, and procedural memory (Squire & Wixted, 2011). Procedural memory is involved in the acquisition of skills and habits (Frith & Frith, 2012; Kaufman et al., 2010; Ullman, 2004) and this memory system is the main focus of the dissertation. Non-declarative memory is traditionally defined as a system that lacks consciousness and lacks dependence on the medial temporal lobe (Reber, 1967). Memory systems can also be differentiated based on the consciousness of

the memory representations. Based on this approach, explicit and implicit memory systems can be distinguished (Graf & Schacter, 1985), where the former refers to consciously accessible memories and the latter refers to a system that lacks consciousness. These partly overlap with the declarative/non-declarative distinction: declarative memories are consciously accessible, that is, explicit, whereas non-declarative memories are implicit, that is, consciously inaccessible (Reber, 2013). These terms are therefore often used interchangeably.

When investigating human memory, besides differentiating between the memory systems, we also have to take into consideration the different phases of memory. The first phase is learning or acquiring information, followed by consolidation. Through consolidation, the fragile and unstable encoded memory representations became more stable and less susceptible to future interferences, ensuring that the memory representations can be retrieved later (Walker, 2005). Here, we define successful consolidation in the following way: it can manifest either as retention, which refers to similar performance at the end of learning and during subsequent testing or as offline gains, which refers to better performance during testing than learning (termed offline learning, Robertson, Pascual-Leone, & Miall, 2004). The present dissertation investigates both learning and consolidation within procedural memory.

2. Different regularities within procedural memory

In my dissertation, I will focus on procedural memory. Procedural memory gives us the ability to detect and extract many types of regularities from the environment, enabling us to adapt to our surroundings. When humans are exposed to structured patterns, they can acquire the regularities underlying the structured pattern and it typically occurs incidentally (i.e., without the intention to learn) and without awareness (Batterink, Paller, & Reber, 2019; Conway, 2020; Siegelman, Bogaerts, Christiansen, & Frost, 2017). Procedural memory is a multifaceted system which supports several automatic functions, such as learning sequences or probabilistic categorization, and it also underlies the acquisition of numerous motor and cognitive skills (i.e., language) and is related to habits (Frith & Frith, 2012; Kaufman et al., 2010; Ullman, 2004).

Humans are capable of extracting several different kinds of regularities. Prior studies have proposed that the regularities, that is, the input structure can vary in complexity and the different regularities can be placed along a continuum from quite simple, deterministic

ones to more complex, probabilistic ones (Conway, 2020). In relation to sequence learning – which is a function supported by procedural memory – Conway and Christiansen (2001) suggested three main types of sequences. The simplest ones are fixed sequences, where elements occur in the same deterministic, inflexible, and arbitrary order, such as stereotyped sequence of actions. In language, examples for fixed sequences are idioms and stock phrases, such as 'once upon a time'. The second type is statistical-based sequences, where certain pairs or triplets of elements frequently co-occur as defined by the transitional probabilities between the elements. In language, transitional probabilities are crucial in word segmentation (Saffran, Aslin, & Newport, 1996). The third, most complex type of sequences are hierarchical-based structures, where lower-level units are combined into more complex units, just as words are combined into phrases and sentences in language. Relatedly, Dehaene, Meyniel, Wacongne, Wang, and Pallier (2015) proposed a taxonomy where they differentiated between five possible, distinct levels of abstraction when learning a sequence. Humans can learn the transitions from one item to the next and timing. Transitions can correspond to the statistical-based sequences in the taxonomy of Conway and Christiansen (2001). The second level, that is, chunking refers to the mechanisms where we group together recurring elements and store them as a single unit. The third level is ordinal knowledge, which is an ordered list where we have explicit knowledge of which element comes first, second and so on. This level can be related to fixed sequences in the taxonomy of Conway and Christiansen (2001). The fourth and five levels are algebraic patterns and nested tree structures. Moreover, in his review, Conway (2020) further detailed the types of regularities and proposed a rough taxonomy of input structures. The simplest ones are (1) repetitions, followed by (2) serial transitions and (3) chunks. More complex patterns would be (4) finite state grammars, (5) non-adjacent dependencies and (6) hierarchical structure at the top. All of these taxonomies point towards the same direction: the regularities humans can acquire might be placed along a spectrum based on their complexity (Conway, 2020).

Most of the prior studies investigated only one regularity at a time. In my dissertation, I will focus on two kinds of regularities, namely statistical, probability-based regularities, and sequential, serial order-based, rule-based regularities and examine them simultaneously within one paradigm. In the present framework, *statistical / probability-based regularities* refer to shorter-range relations between stimuli based primarily on probabilistic information. The probability-based information is typically learned relatively rapidly and in an incidental manner. The acquisition of *serial order-based*

regularities refers to the learning a series of repeating elements which occur in the same order, typically with embedded noise between them. Probability-based regularities would correspond to the statistical-based sequences in the taxonomy of Conway and Christiansen (2001) and to the transitions in the taxonomy of Dehaene et al. (2015), while serial order-based regularities would correspond to fixed sequences in the taxonomy of Conway and Christiansen (2001) and to ordinal knowledge in the taxonomy of Dehaene et al. (2015).

In this section, I will present the empirical evidence supporting the distinction of probability-based and serial order-based regularities. Previous studies have shown that the acquisition of these regularities differs both on the behavioral and neural level. On the behavioral level, the learning of serial order-based regularities seems to follow a gradually increasing trajectory, whereas the acquisition of probability-based regularities seems to plateau early (Kóbor et al., 2018; Simor et al., 2019). The distinct learning trajectories of the two regularities were also reflected in event-related potentials (Kóbor et al., 2018). In detail, the N2 component reflected the acquisition of probability-based information already at the beginning of the task and it did not change with practice, thus, it showed the rapid, automatic detection of probability-based regularities. The acquisition of serial order-based information was mirrored by the N2 and P3 components, both gradually changed as the task progressed and learning took place. Relatedly, Takács et al. (2021) demonstrated that the two regularities seem to be associated with distinct aspects of information coded in the N2 time window. Moreover, Nemeth, Janacsek, and Fiser (2013) examined the developmental trajectory of the two regularities from the age of 11 to 39 years and showed age-invariant trajectory regarding the acquisition of both probability-based and serial order-based regularities. Furthermore, Simor et al. (2019) found differences in neural oscillations during consolidation of the two regularities. Further improvements in the learning of serial order-based information were predicted by slow frequency oscillations (high delta and theta power) during sleep, whereas consolidation of probability-based regularities was not associated with spectral EEG power measures. Moreover, the two regularities also seem to have different functional connectivity patterns during consolidation (Zavecz, Janacsek, Simor, Cohen, & Nemeth, 2020). To sum up, converging evidence show differences both in the acquisition and consolidation of probability-based and serial order-based regularities.

3. Measuring regularities within procedural memory with different versions of the Alternating Serial Reaction Time task

To measure the different kinds of regularities within procedural memory, we employed two versions of Alternating Serial Reaction Time (ASRT) task (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013) in all studies. The ASRT task is a visuo-motor four-choice reaction time task. Participants see four horizontal, equally spaced circle on the screen. In one of the circles, a stimulus appears, and participants have to press the corresponding button as fast and as accurately as possible. The task is bimanual, participants use the index and middle finger of both hands.

The appearance of the stimuli follows an 8-element alternating sequence, where pattern and random elements alternate with each other (e.g., 1-r-2-r-4-r-3-r, where numbers represent the locations from left to right and 'r' indicates a randomly selected location out of the four possible ones). Due to the alternating sequence, some runs of three consecutive trials (termed triplets in the present studies) occur with a higher probability (or frequency) than others (for details on how probability/frequency differs between triplets, see the task description in the studies). In the present studies, we used the terms frequency and probability interchangeably as joint probability (how frequently a particular triplet occurs) and conditional probability (how predictable the last element of the triplet is) are identical in the ASRT task in relation to triplets (Szegedi-Hallgató, Janacsek, & Nemeth, 2019). Each trial is categorized as the last element of a triplet in a moving window manner, that is, a given trial is defined as the first element of a given triplet, as the second element of the next triplet and as the last element of the following triplet.

Participants show faster reaction times and higher accuracy to high-probability triplets compared to low-probability ones (e.g., Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013; Song, Howard, & Howard, 2007a, 2007b). Importantly, however, comparing the triplets based only on probability in the original ASRT task has a confound: it does not take into account whether the last trial of the triplet is a pattern or random trial, meaning that it does not take into account the serial-order information (Nemeth, Janacsek, & Fiser, 2013). Hence, probability-based and serial order-based regularities are intermixed in the original ASRT task. As mentioned, besides probability, we can differentiate between the triplets based on their structure, that is, whether they are pattern or random trials (Nemeth, Janacsek, & Fiser, 2013). High-probability triplets can

be separated into pattern-ending high-probability triplets and random-ending high-probability triplets. The third element of a low-probability triplet can only be random as pattern elements always occur with high probability. To sum up, we can differentiate between three trial types: (1) *pattern trials* belong to the predetermined sequence and are the last element of a high-probability triplet, (2) *random high* trials appear randomly and are the last element of a high-probability triplet, and (3) *random low* trials also appear randomly but are the last element of a low-probability triplet.

In the original ASRT task, performance is measured by comparing accuracy scores and reaction times on high-probability (both pattern and random high trials) and low-probability triplets (random low trials). Over the years, the previous studies used different terms to the performance measured by the original ASRT task: probabilistic sequence learning and statistical learning are the most common ones (e.g., Kóbor, Janacsek, Takács, & Nemeth, 2017; Takács et al., 2018). Here, we employ the original ASRT task in Study 4, where we refer to the measured learning performance as statistical learning, highlighting that the main difference between the compared trials is their probability.

To disentangle statistical / probability-based and serial order-based regularities, we also used a modified version of the task, named cued ASRT task (or explicit ASRT task in some studies, Nemeth, Janacsek, & Fiser, 2013). The underlying structure (i.e., the alternating sequence) is identical in the original and cued ASRT tasks. However, in the cued ASRT task, pattern and random elements are marked by different visual stimuli: pattern elements are indicated by a dog's head and random elements are denoted by a penguin. Participants are informed about the presence of the alternating sequence and about the fact that the occurrence of the dogs follow a predetermined pattern, and the penguins appear in a random location. They are not informed about the exact sequence or the length of the sequence, but they are instructed to discover the pattern of the dogs' appearance to improve their performance. These modifications enable us to measure the acquisition of probability-based and serial order-based regularities in parallel, within one learning session.

In the cued ASRT task, the acquisition of probability-based regularities is measured by the difference in accuracy scores or reaction times between random high and random low trials. These trials share the sequence (serial-order) properties as they both random trials, but they differ in probabilistic (statistical) properties, as random high trials are part of a high-probability triplet and random low trials are part of a low-probability triplet. The acquisition of serial order-based regularities is measured by the difference in

accuracy scores or reaction times between pattern and random high trials. These trials are both high-probability trials, hence, they share the same probabilistic (statistical) properties, however, they differ in sequence (serial-order) properties as pattern trials are part of the predetermined sequence while the others appear randomly (Nemeth, Janacsek, & Fiser, 2013). In our previous studies (e.g., Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019) and in Study 1 and 3, we also refer to these learning processes as statistical and sequence learning, respectively.

Importantly, these two regularities can be distinguished even in the original, non-cued version of the ASRT task, if participants are exposed to enough learning sessions. Howard and Howard (1997) found that young adults show better performance on pattern trials compared to random high-probability trials after extensive learning (after a 4-day learning with 21 80-trial blocks in each day). Similarly, Howard, Howard, Japikse, DiYanni, et al. (2004) showed that young adult participants separated pattern trials from random high-probability trials after excessive training (after a 9-day learning with 21 80-trial blocks in each day). These studies used the original ASRT task, where both pattern and random elements are marked with the same visual stimuli. Prior studies showed that a cued, instructed version of the task (i.e., where pattern and random stimuli are marked with different visual stimuli and participants are instructed to find the sequence) results in relatively faster acquisition of the alternating sequence (i.e., within one learning session; e.g., Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019). Therefore, acquisition of serial order-based regularities (i.e., differentiating between pattern and random high trials) can happen both in incidental learning condition (original, non-cued ASRT task, Howard, Howard, Japikse, DiYanni, et al., 2004; Howard, & Howard, 1997) and intentional learning condition (cued ASRT task, Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019). Intentional learning condition typically leads to faster learning of serial order-based regularities than incidental learning condition.

Besides the regularities, other learning processes can also be present in the ASRT task. One such learning process is called general skill learning, which means that as a result of practice, participants process and respond faster to stimuli, and show improved visuomotor coordination, independent of the regularities embedded in the visual stimulus stream (Hallgato, Győri-Dani, Pekár, Janacsek, & Nemeth, 2013; Juhasz, Nemeth, & Janacsek, 2019). Out of the four studies, we examined general skill learning only in Study

4, where we describe the lifespan trajectory of the consolidation of statistical as well as general skill knowledge.

In three studies out of the four presented in the dissertation, we employed the cued ASRT task to investigate the acquisition and consolidation of both probability-based and serial order-based regularities. In the published studies, we used different names when describing these regularities. Statistical regularities have been referred to as probability-based regularities as well as statistical learning, whereas serial order-based regularities have been referred to as sequential regularities and as sequence learning. In the general introduction, we consistently use the terms probability-based and serial order-based regularities. In Study 4, we used the original ASRT task to investigate the consolidation of statistical regularities and general skill knowledge. We decided to refer to the former learning performance as statistical learning to highlight that the main difference between the compared trials in the original ASRT task is their probability. Furthermore, a prior study has shown that both the acquisition and consolidation of statistical regularities are comparable in the original and cued ASRT task (Horvath, Torok, Pesthy, Nemeth, & Janacsek, 2020). Participants who were exposed to the statistical regularities using the original ASRT task showed comparable performance to those who were exposed to identical regularities using the cued ASRT task. Nevertheless, statistical regularities in the original and cued ASRT task differ on the operational level as in the original task, serial-order information is intermixed with the statistical information. Based on this and for an easier understanding, we decided to use different terms for statistical, probability-based regularities in the original and cued ASRT tasks. When referring to the regularity in the original ASRT task, we use the term statistical regularity, and in case of the cued ASRT task, we use the term probability-based regularity. For further details of the characteristics of the ASRT tasks, see the task descriptions in the respective studies.

4. Procedural memory in Tourette Syndrome

Investigating procedural memory is highly relevant both in typical and atypical development. Considering the latter, in my dissertation, I will focus on a childhood-onset neurodevelopmental disorder, namely Tourette Syndrome (TS, or Tourette disorder). According to DSM-V, the diagnosis requires several motor tics and at least one vocal tic to be present for at least a one-year period with an onset before the age of 18, and tics cannot be accounted for other medical conditions or medication (APA, 2013). Tics are

recurrent, abrupt, semi-voluntary movements or vocalizations, which are typically mimic some fragment of normal behavior, but they are misplaced in context and time. Simple and complex tics are distinguished: the former ones seem like purposeless movements or vocalizations and involve only a few muscles, such as blinking, shrugging, coughing or throat clearing. Complex tics can appear as more purposeful, goal-directed movements or sounds, they are more coordinated, such as spinning, hopping, or repeating sounds made by oneself (palilalia) or by others (echolalia). Boys are more frequently affected than girls (3:1 ratio) and the prevalence rate is around 1% (Robertson, 2015). Typically, the first tics appear around the age of 4-6 years, the highest severity is around the age of 12 years, followed by a substantial decline during adolescence. Around 20% of the patients continue ticcing during adulthood as well (Bloch & Leckman, 2009; Robertson et al., 2017). Tics change over time with some tics disappearing and new tics appearing, and they also show considerable fluctuations in frequency and distribution and typically wax and wane in severity.

As of right now, there is no generally accepted theoretical model describing the development of tics. Prior studies suggested, however, a potential link between tics and habits (Conceição, Dias, Farinha, & Maia, 2017; Goodman, Marsh, Peterson, & Packard, 2014; Maia & Conceição, 2017). In detail, tics resemble habits, and evidence for this link can be found both on the behavioral and neural levels. On the behavioral level, both tics and habits are automatically executed, inflexible responses to stimuli that are hard to inhibit. Considering the neural level, TS is characterized by structural and functional alterations in the basal ganglia, in the related frontal regions and in the cortico-basal ganglia-thalamo-cortical (CBGTC) circuits. Tic-related activation has been found in the premotor cortex, sensorimotor cortex, supplementary motor area, putamen, globus pallidus and thalamus (Bohlhalter et al., 2006; Stern et al., 2000; Wang et al., 2011; Worbe et al., 2015). Moreover, motor tic severity was positively correlated with activation within the anterior mid-cingulate cortex and posterior cingulate cortex (Jackson, Sigurdsson, Dyke, Condon, & Jackson, 2021). Increased volume of the putamen (Roessner et al., 2011) and decreased volume of the caudate nucleus (Peterson et al., 1998; Peterson et al., 2003) has also been shown in TS; and the structural connectivity within the motor loop and resting-state functional connectivity in the motor networks are also increased in TS (Worbe et al., 2012; Worbe et al., 2015). TS also involves striatal dopaminergic hyperinnervation (e.g., Buse, Schoenefeld, Münchau, & Roessner, 2013; Singer, 2013) and Maia and Conceição (2017) proposed that tics are

maladaptive motor habits reinforced by increased phasic dopamine responses, while tonic dopamine responses promote the execution of the learned tics. Besides being altered in TS, these brain areas are functionally related to the formation of skills and habits in the procedural memory as well. Indeed, procedural memory has been linked to the basal ganglia and CBGTC circuits (Doyon et al., 2009; Janacsek et al., 2020; Poldrack & Packard, 2003). To sum up, tics and habit are similar on the behavioral level and are related to similar neural networks. These shared similarities make the investigation of procedural memory in TS highly relevant.

In Study 1 and 2, I will focus on the possible alteration of procedural memory in TS. Prior studies proposed that the neural alterations in TS lead to a hyperfunctioning of the cognitive functions related to the altered brain regions (Clinical Extension Hypothesis, Dye, Walenski, Mostofsky, & Ullman, 2016). As detailed above, one cognitive function which shares a neural overlap with the neural background of TS is procedural memory. Previous studies have already suggested that enhanced procedural functions, that is, procedural hyperfunctioning can be present in the disorder (Dye et al., 2016; Shephard, Groom, & Jackson, 2019; Takács et al., 2018; Walenski, Mostofsky, & Ullman, 2007). In Study 1, we investigated the acquisition of procedural information, especially focusing on which regularities (probability-based or serial order-based) related to procedural memory might be altered in TS. In Study 2, instead of learning, we examined the consolidation of probability-based and serial order-based regularities, following both short-term and long-term offline delays. Altogether, these two studies will provide information about how the functional changes in fronto-striatal circuits in atypical development influence procedural memory. As these circuits are related to both procedural learning and consolidation (Doyon et al., 2009), we can expect changes in both phases.

5. Procedural memory in typical development

The fronto-striatal circuits as well as the hippocampus undergo a protracted development, therefore, investigating how procedural memory develops is a crucial question not only in atypically developing disorders where these brain areas are involved, but in typical development as well. The developmental trajectory of procedural learning has been described with three different models (for a review, see Zwart, Vissers, Kessels, & Maes, 2019), whereas, to the best of our knowledge, no theoretical model has been

proposed for the developmental trajectory of procedural memory consolidation. In the next section, I will present the models describing the lifespan trajectory of procedural learning and shed light on the need for examining procedural memory consolidation across the lifespan.

The first model for the lifespan trajectory of procedural learning proposed age invariance (Reber, 1993) and was based on studies finding similar performance in children and adults (e.g., Meulemans, Van der Linden, & Perruchet, 1998) and studies showing that procedural learning is related to the striatum, which matures early in life (Reber, 1993). This model suggests that no developmental change occurs from childhood to adulthood. In contrast, the other two models propose age variance in procedural learning. The inverted U-shaped model describes the following trajectory: procedural learning undergoes a gradual improvement during childhood and adolescence, peak performance is present in young adulthood and a decline emerges with aging (Lukács & Kemény, 2015). A study comparing learning from the age of 7 to 87 years and showing better performance in young adulthood compared to childhood and old adulthood supports this model (e.g., Lukács & Kemény, 2015). The other age-variant model, which can be referred to as the 'competition model' proposes a different trajectory: better procedural learning in childhood (under the age of 12) is followed by less effective learning later in life (Janacsek, Fiser, & Nemeth, 2012). This model is supported by empirical studies employing a lifespan approach and showing better learning under the age of 12 (Janacsek et al., 2012; Juhasz et al., 2019; Nemeth, Janacsek, & Fiser, 2013). Out of the three competing models, the age-variant models are strongly supported by empirical findings, however, the exact trajectory is not full apparent yet. The contradictory results of the two age-variant models might be due to the methodological differences between them (Zwart et al., 2019).

Considering the developmental curves for procedural learning, different trajectories can be proposed for the consolidation of procedural knowledge. The age-variant trajectories for procedural learning raises the question whether the consolidation of such knowledge might follow an age-variant curve as well. However, this is not necessarily the case. In atypical development, a dissociation between learning and consolidation has been demonstrated: one study showed enhanced learning and intact consolidation in Tourette syndrome (Takács et al., 2018), while in another study, intact learning was followed by impaired consolidation in developmental dyslexia (Hedenius, Lum, & Bölte, 2021). Nevertheless, it is still an open question whether this dissociation can be found in

neurotypical population as well. Hence, in Study 3 and 4, we concentrated on the consolidation of procedural knowledge. In more detail, Study 3 investigated the long-term, that is, one-year consolidation of probability-based and serial order-based regularities in childhood and adolescence. In Study 4, we aimed to examine the lifespan trajectory of statistical and general skill knowledge involving participants from the age of 7 to 76 years in a cross-sectional design.

6. Research questions

6.1. Is the learning of probability-based and serial order-based regularities enhanced in Tourette syndrome?

As detailed above, examining procedural memory in TS is highly relevant due to the behavioral and neural overlap between this memory system and tics. Based on the prior empirical studies, the picture is a bit mixed: studies mostly showed enhanced (Dye et al., 2016; Shephard et al., 2019; Takács et al., 2018; Walenski et al., 2007) or at least intact (Channon, Pratt, & Robertson, 2003; Takács et al., 2017) procedural memory performance in the disorder, but impaired performance has also been suggested (Keri, Szlobodnyik, Benedek, Janka, & Gadoros, 2002; Marsh et al., 2004). In Study 1, we investigated how two regularities within the procedural memory system, namely probability-based and serial order-based regularities are affected in TS and whether these regularities contribute to the possible procedural hyperfunctioning suggested by prior studies (Dye et al., 2016; Shephard et al., 2019; Takács et al., 2018; Walenski et al., 2007).

6.2. Is the consolidation of probability-based and serial order-based regularities robust in Tourette syndrome?

In Study 2, we focused on the consolidation of procedural knowledge in TS which has not received much attention so far. Takács et al. (2018) used the original ASRT task and employed a 16-hour offline delay in their study design. Their results showed enhanced learning in the disorder, however, strong conclusions could not be drawn regarding consolidation. The TS group showed greater forgetting than the typically developing control groups, but this could be explained by the learning differences at the end of the learning phase. When they compared the overnight changes as a function of prior knowledge – meaning that they controlled for the learning differences – the groups showed similar retention performance. Consolidation of procedural knowledge has also

been examined indirectly in TS by investigating the access to previously consolidated procedural knowledge using language-based tasks. Both studies (Dye et al., 2016; Walenski et al., 2007) showed evidence for faster access, which falls in line with the notion of procedural hyperfunctioning in TS. Here, we examined both the short-term (five-hour) and long-term (one-year) consolidation of probability-based and serial order-based regularities in TS. Employing a one-year offline period enables us to bring closer the time scale of lab studies, which typically last only for hours or days and of real-world observations, that is, the time scale of learning a new skill or develop a habit in everyday life.

6.3. Is the one-year consolidation of probability-based and serial order-based regularities successful in typically developing children and adolescents?

In Study 3, we recruited participants between the ages of 9 and 15 years and focused especially on the long-term, that is, one-year consolidation of procedural knowledge. Procedural memory performance over short periods of time, that is, from minutes to weeks has been investigated across the lifespan. Long-term memory performance, that is, from months to years, received relatively less attention and has not been assessed in children and adolescents yet. In neurotypical adults, both Romano, Howard, and Howard (2010) and Kóbor et al. (2017) showed robust representation of statistical knowledge after a one-year offline delay. Several studies showed retained knowledge in typically developing children following various offline delays (from 11-hour to one week, e.g., Fischer, Wilhelm, & Born, 2007; Hedenius et al., 2021; Hedenius et al., 2013; Hedenius et al., 2011; Nemeth, Janacsek, Balogh, et al., 2010; Takács et al., 2018), but no studies tested consolidation by employing a longer offline period. As mentioned in the previous section, the possible dissociation of learning and consolidation of procedural knowledge raises the question whether, similarly to age-variant learning, consolidation of procedural knowledge might also be age-variant. Hence, in Study 3, we investigated the long-term, that is, one-year consolidation of probability-based as well as serial order-based knowledge in a sample of 9-15-year-olds.

6.4. Does the consolidation of procedural knowledge follow an age-variant trajectory across the lifespan?

In Study 4, we aimed to unveil the lifespan differences in the consolidation of procedural knowledge by employing the same experimental design in a large sample of participants between the ages of 7 and 76 years. Here, we examined both statistical and general skill knowledge. Regarding both types of knowledge, prior studies mostly focused on one age group at a time or contrasted only a few age groups. Most studies have suggested successful retention of statistical knowledge in children and adolescents (e.g., Hedenius et al., 2021; Hedenius et al., 2013; Hedenius et al., 2011; Nemeth, Janacsek, Balogh, et al., 2010; Takács et al., 2018), and in young and middle-aged adults (e.g., Arciuli & Simpson, 2012; Horvath et al., 2020; Kim, Seitz, Feenstra, & Shams, 2009; Kóbor et al., 2017; Meier & Cock, 2014; Nemeth, Janacsek, Király, et al., 2013; Romano et al., 2010; Simor et al., 2019; Song et al., 2007b), but studies involving older adults have shown mixed results: some have found intact retention (Romano et al., 2010), and others have suggested a decline (Nemeth & Janacsek, 2011).

As for the consolidation of general skill knowledge, offline improvement (i.e., better performance in the testing phase compared to the end of the learning phase) has been found in children and adults (Csábi, Benedek, Janacsek, Katona, & Nemeth, 2013; Csabi et al., 2015; Hedenius et al., 2021; Hedenius et al., 2013; Nemeth, Janacsek, Balogh, et al., 2010; Nemeth, Janacsek, Király, et al., 2013). However, we have no information about whether the magnitude of offline improvement differs between children and adults. In older adults, once again, the picture is not clear: some studies have found offline improvement in older adults, although the gain was smaller compared to the gain in young adults (Nemeth, Janacsek, Londe, et al., 2010), while other studies could not detect offline learning in older adults (Nemeth & Janacsek, 2011). Based on these empirical findings, no strong conclusions can be drawn either for statistical or general skill knowledge. For statistical knowledge, retention in all age groups seems to be the most plausible outcome, which would support an age-invariant model of the consolidation of statistical knowledge. For general skill knowledge, offline improvement could be expected in most age groups, the extent of improvement could differ.

To sum up, the aim of the studies in the dissertation is to investigate how the changes in fronto-striatal networks might lead to changes in procedural memory and to provide a deeper understanding of this memory system both in typical and atypical

development. In typical development, the aim is to shed light on the developmental trajectory of procedural memory consolidation over various offline delays. In atypical development, our aim is to reveal which processes within the procedural memory system are altered in a fronto-striatal disorder, namely in Tourette syndrome. Exploring the impairments and strengths in a disorder could serve as a basis for developing new training approaches which may help reduce some disadvantages associated with the disorder. The main characteristics and questions of the four studies are summarized in Table 1.1.

Table 1.1. Summary of the studies presented in the dissertation.

	Population	**Measured processes**	**Task**	**Main question**
Study 1	children and adolescents with TS and matched TD controls	acquisition of probability-based and serial-order based regularities	cued ASRT task	Is the learning of probability-based and serial order-based regularities enhanced in Tourette syndrome?
Study 2	children and adolescents with TS and matched TD controls	consolidation of probability-based and serial order-based regularities	cued ASRT task	Is the short- and long-term consolidation of probability-based and serial-order regularities robust in Tourette syndrome?
Study 3	TD children and adolescents	consolidation of probability-based and serial order-based regularities	cued ASRT task	Can children and adolescents successfully retain probability-based and serial-order regularities over a one-year delay?
Study 4	TD children and adolescents, neurotypical young, middle-aged, and old adults	consolidation of statistical and general skill knowledge	original ASRT task	Are the lifespan trajectories of statistical and general skill knowledge age-variant?

Note. TS = Tourette syndrome, TD = typically developing, ASRT task = Alternating Serial Reaction Time task.

II. Study 1: Dissociation between two aspects of procedural learning in Tourette syndrome: Enhanced statistical and impaired sequence learning

Publication:

Tóth-Fáber, E., Tárnok, Z., Janacsek, K., Kóbor, A., Nagy, P., Farkas, B. C., Oláh, Sz., Merkl, D., Hegedűs, O., Nemeth, D., & Takács, Á. (2021). Dissociation between two aspects of procedural learning in Tourette syndrome: Enhanced statistical and impaired sequence learning. *Child Neuropsychology*, 27(6) 799-821.
https://doi.org/10.1080/09297049.2021.1894110

Abstract

Tourette syndrome (TS) is a childhood-onset neurodevelopmental disorder that primarily affects the cortico-basal ganglia-thalamo-cortical (CBGTC) circuitry and is characterized by motor and vocal tics. Previous studies have found enhancement in procedural memory, which depends on the CBGTC circuitry and plays an important role in the learning and processing of numerous motor, social, and cognitive skills and habits. Based on these studies, procedural hyperfunctioning in TS has been proposed. However, the neurocognitive mechanism underlying such hyperfunctioning is poorly understood. Here, we investigated how two aspects of procedural learning, namely 1) frequency-based statistical learning and 2) order-based sequence learning, are affected in TS. Twenty-one children with TS between the ages of ten and fifteen as well as 21 typically developing controls were tested on a probabilistic sequence learning task that enables the parallel assessment of these two aspects. We found that children with TS showed enhanced sensitivity to statistical information but impaired sequence learning compared to typically developing children. The deconstruction of procedural memory suggests that procedural hyperfunctioning in TS may be supported by enhanced sensitivity to statistical information. These results can provide a potential path for improving therapy methods and skill-oriented educational programs for TS.

Keywords: Tourette syndrome, sub-cortical structures, procedural memory, skill learning, basal ganglia, statistical learning

1. Introduction

Tourette syndrome (TS) or Tourette Disorder is a neurodevelopmental disorder characterized by several motor and at least one vocal tic for at least a one-year period, which cannot be accounted for other medical conditions or medications (APA, 2013). TS can be marked by altered cognitive functions, including both impairments and enhancements (e.g., Delorme et al., 2016; Jung, Jackson, Nam, Hollis, & Jackson, 2015; S. C. Mueller, Jackson, Ranu, Sophia, & Hollis, 2006; Palminteri et al., 2011; Takács et al., 2018; Yaniv et al., 2017, although comorbidities could have confounded some of the prior results see Morand-Beaulieu et al., 2017). One of the areas where such enhancement has been found is procedural learning (and also accessing the already established procedural information), which underlies the learning of motor and cognitive skills (Delorme et al., 2016; Dye et al., 2016; Palminteri et al., 2011; Takács et al., 2018; Walenski et al., 2007). Based on these studies, it has been proposed that procedural hyperfunctioning exists in TS. The overlap between the neurobiological characteristics of TS and the neurophysiological underpinnings of procedural memory provides an opportunity to examine (1) the aspects of procedural functions and (2) the cognitive models of TS (Dye et al., 2016; Shephard et al., 2019) at the same time. Here we aimed to examine how two aspects of procedural learning, namely statistical and sequence learning, are affected in TS and test whether these aspects contribute to the procedural hyperfunctioning proposed by previous studies.

The cognitive profile of TS has been thoroughly investigated, including executive functions (Channon et al., 2009; Yaniv et al., 2017), social cognition (Eddy & Cavanna, 2013), and procedural memory (Channon et al., 2003; Marsh et al., 2004). Results on the cognitive profile could be confounded by certain factors such as comorbidities (for a review, see Morand-Beaulieu et al., 2017). A crucial aspect of cognition is procedural memory, which contributes to the acquisition, storage, and use of implicit motor and automatic cognitive behaviors such as skills and habits (Kaufman et al., 2010; Ullman, 2004). Procedural memory is mediated mainly by the basal ganglia, particularly the striatum, and relies on the cortico-basal ganglia-thalamo-cortical (CBGTC) pathways (Doyon et al., 2009; Janacsek et al., 2020; Poldrack & Packard, 2003). The basal ganglia are thought to contribute to the acquisition of skills and habits, whereas neocortical regions might be more important for processing skills after they have been automatized (Ullman, 2016). In TS, tics and habits are phenomenologically similar and share neural

underpinnings (Conceição et al., 2017). It has been suggested that alterations in the frontostriatal regions and improper procedural learning mechanisms can explain the hyperkinetic profile of TS (Albin & Mink, 2006).

The exact brain mechanisms underlying TS are not yet fully understood. However, converging evidence suggests both structural and functional abnormalities in the basal ganglia, related frontal regions, and in the CBGTC pathways (Albin et al., 2003; Albin & Mink, 2006; Maia & Frank, 2011; Mink, 2001; Peterson et al., 1998; Peterson et al., 2003; Stern et al., 2000). Tics may reflect abnormal habit-learning mechanisms, where improper stimulus-response associations are learned (Albin & Mink, 2006; J. Goodman et al., 2014; Petruo et al., 2019). Abnormalities in the CBGTC loop support the hypothesis of altered habit-learning in TS. Tics may result from a heightened direct pathway activity relative to indirect pathway activity in the CBGTC loop (Maia & Frank, 2011; Mink, 2001).

These neurobiological alterations may lead not only to tics but also to enhancements in procedural learning (Dye et al., 2016; Walenski et al., 2007). Most of the previous studies examining procedural learning in TS reported enhanced functions (Delorme et al., 2016; Dye et al., 2016; Palminteri et al., 2011; Shephard et al., 2019; Takács et al., 2018; Walenski et al., 2007), or at least intact functions (Channon et al., 2003; Takács et al., 2017), with only two reporting impaired performance (Keri et al., 2002; Marsh et al., 2004). The reason for the differences among these studies are not yet clear. The reason for the different results among these studies is not yet clear. Some prior studies involved only a handful of TS participants, which could have led to low statistical power failing to find group differences. Previous studies also diverse in terms of age (child or adult TS samples) and in terms of tic severity; both might be differently related to procedural learning. Relatedly, the presence of comorbidities could also confound the results. Another possibility is that previous studies tapped into different aspects of procedural memory and these aspects are differentially affected by TS. In the present study, we focused on this and investigated two aspects of procedural learning in TS.

Previous studies showing intact or enhanced procedural learning measured either sequence learning or language performance. One of the first studies using a sequence learning task to measure procedural memory in TS found no group differences between the TS and control groups (Channon et al., 2003). Similarly, Takács et al. (2017) reported comparable learning performance between children with TS and typically developing peers using the Alternating Serial Reaction Time (ASRT) task. Shephard et al. (2019) examined sequence learning using the Serial Reaction Time (SRT) task in children with

TS. Participants were assessed on two types of blocks: (1) sequence blocks containing stimuli following a predetermined sequence and (2) non-sequence blocks with random stimuli. Children with TS showed difficulties transitioning from sequence to non-sequence blocks, they showed greater disruption in accuracy compared to the control group. This result can imply that children with TS overlearned the sequence in the task, which led to a more difficult transition. That is, children with TS showed procedural hyperfunctioning. In support of this, Takács et al. (2018) reported evidence of enhanced procedural learning in TS using the ASRT task. Children with TS made more prediction errors through learning than their typically developing peers, indicating enhanced sensitivity to the underlying regularities of the task. Moreover, procedural memory performance in TS had an early peak, and typically developing (TD) children did not exceed the level of TS performance throughout the task. The notion of procedural hyperfunctioning is further supported by studies with adult TS population, as well (Delorme et al., 2016; Palminteri et al., 2011). Palminteri et al. (2011) found that adults with TS showed enhanced reinforcement learning in a motor learning task. Furthermore, Delorme et al. (2016) found a higher rate of response to previously learned but devalued stimulus-response-outcome associations, which also suggest enhanced procedural functions.

Similar to the sequence learning results, procedural hyperfunctioning has also been found in language-based tasks. Two studies (Dye et al., 2016; Walenski et al., 2007) showed faster grammatical processes in TS on the morphological and phonological levels. Walenski et al. (2007) was the first study to demonstrate enhanced procedural functions in TS. In this study, children with TS showed faster producing of rule-governed past tenses compared to their typically developing peers (e.g., slip-slipped) while showing similar performance on producing irregular past tenses (e.g., bring-brought). Moreover, the naming of manipulated objects (e.g., hammer) was also "speeded" in TS, while naming of non-manipulated objects (e.g., elephant) was similar in the TS and control groups. These results support the procedural hyperfunctioning hypothesis in TS. Whereas producing regular past tenses and naming manipulated objects both rely on procedural memory, producing irregular past tenses and naming non-manipulated objects appear to be stored in declarative memory (Ullman, 2004). Another language-related study (Dye et al., 2016) further strengthens procedural hyperfunctioning in TS. This study showed "speeded" grammatical composition on a non-word repetition task in TS. Children with TS repeated non-words (e.g., "naichovabe") faster than typically developing peers, while

in accuracy there was no difference between the groups. This type of phonological manipulation taps into decomposition, a procedural aspect of the language domain.

Procedural memory is a complex system and it supports several functions, such as learning sequences, probabilistic classification, and aspects of language, including grammar (Fiser & Aslin, 2001; Howard, & Howard, 1997; Knowlton, Squire, & Gluck, 1994; Ullman, Earle, Walenski, & Janacsek, 2020). Converging evidence suggests procedural hyperfunctioning in TS in both the acquisition of procedural information (such as in sequence learning; Delorme et al., 2016; Palminteri et al., 2011; Takács et al., 2018) and the accessing of the already established procedures (such as in grammatical processing; Dye et al., 2016; Walenski et al., 2007). Importantly, acquiring procedural information is a complex function relying on multiple, parallel learning processes (Maheu, Dehaene, & Meyniel, 2019; Maheu, Meyniel, & Dehaene, 2020; Siegelman et al., 2017; Thiessen, Kronstein, & Hufnagle, 2013). It is not yet clear which aspect of procedural learning supports the potential hyperfunctioning in TS.

Based on the previous studies, it is still unclear which aspects of procedural learning are affected in TS. Dye et al. (2016) suggest the importance of processing sequential information. Children and adults with TS may have enhanced sequence sensitivity, which leads to enhanced sequence learning and grammatical processing. Another significant aspect of procedural learning is processing of probabilistic information. The results of Takács et al. (2018), where children with TS showed enhanced learning on a probabilistic sequence learning task, suggest that enhanced sensitivity to probabilistic information may contribute to procedural hyperfunctioning. However, neither of these studies focused on contrasting these two aspects of procedural learning. Here, we designed a study to investigate how sensitivity to sequential vs. probabilistic information is affected in TS.

Crucially, the sensitivity to sequential information and to probabilistic information cannot be measured at the same time with most tasks. There is a paradigm, however, designed to distinguish these two learning processes. The cued version of the ASRT task (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013) is able to measure both learning processes in parallel. Here, *statistical learning* refers to the acquisition of probabilistic (frequency) information. Participants learn the shorter-range relationship between visual stimuli that is primarily based on frequency (differentiating between more frequent and less frequent stimulus chunks). Additionally, *sequence learning* refers to the acquisition of order-based information. Thus, participants learn a

series of stimuli that repeatedly occur in the same (deterministic) order, intermixed with random stimuli (resulting in an alternating sequence structure). From a theoretical perspective, it is important to note that at the level of transitional probabilities, statistical learning and sequence learning can be considered similar. Whereas statistical learning (as measured in the ASRT task) refers to the acquisition of second-order transitional probabilities that are less than one, sequence learning refers to the acquisition of second-order transitional probabilities that are equal to one. Despite the fact that both can be viewed as acquisition of transitional probabilities, a growing body of evidence suggests that they exhibit at least partially different characteristics both at behavioral and neural level (Howard, & Howard, 1997; Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019). Shorter-range probabilistic information (i.e., *statistical learning*) is typically acquired incidentally and relatively rapidly (Kóbor et al., 2018; Simor et al., 2019). In contrast, acquisition of the alternating sequence may occur either incidentally or intentionally with gradually improving performance in both cases (although an intentional learning condition typically results in faster sequence acquisition compared to an incidental learning condition) (Howard, Howard, Japikse, DiYanni, et al., 2004; Howard, & Howard, 1997; Simor et al., 2019). Furthermore, statistical and sequence learning appear to have different electrophysiological characteristics, suggested by event-related potentials during learning (Kóbor et al., 2018) as well as by neural oscillations during consolidation (Simor et al., 2019).

The present study focuses on testing procedural hyperfunctioning in children with TS and investigates two aspects of procedural learning, namely, statistical learning and sequence learning using the ASRT task. Since previous ASRT studies showed that a cued, instructed version of the task results in relatively faster acquisition of the alternating sequence (e.g., Kóbor et al., 2018; Simor et al., 2019), enabling to measure statistical and sequence learning in the same time frame (i.e., within one learning session), we also chose this cued version in the current study. As previous research on TS and procedural learning did not provide first-hand evidence on these two aspects of procedural learning, we follow an exploratory approach to test which aspect is affected or may even support the procedural hyperfunctioning in TS.

2. Materials and methods

2.1. Participants

Twenty-seven children diagnosed with TS between the age of 10 and 15 were recruited through a regional child psychiatry hospital in Budapest, Hungary. They had been diagnosed with TS based on the DSM-V criteria (APA, 2013). TS and any comorbidities have been diagnosed by a team of child psychiatrist, clinical psychologist and special education teacher following a minimum one-week-long stay and observation in the hospital. The TS children visit the hospital regularly later as well for check-ups and treatment. Hence, comorbidities have not been evaluated as a part of the present study but were previously diagnosed in the hospital. Children with comorbid psychiatric or neurodevelopmental disorders were excluded from the analysis except for the ones with comorbid attention deficit hyperactivity disorder (ADHD) or obsessive-compulsive disorder (OCD) since the presence of these disorders are common in TS (Robertson, 2015). Three children were excluded from the analyses due to comorbid disorders: one child had comorbid depression, OCD, and ADHD, one child had comorbid depression, anxiety disorder, and ADHD, and one child had comorbid autism spectrum disorder and ADHD. Moreover, medication was also an exclusionary criterion. From the recruited 27 participants, five participants were taking medication during the time of testing. Out of the five participants, two of them had some comorbid diagnoses other than ADHD or OCD. Therefore, additional three participants were excluded from the analyses due to medication. In sum, we excluded six participants due to comorbid diagnoses and medication and data of 21 children with TS (18 boys and three girls) were analyzed. In this final sample, three children met the criteria for comorbid ADHD and one child was diagnosed with comorbid OCD and ADHD.

Ninety-nine typically developing (TD) children participated in the study from local schools. From the TD group, we matched 21 children one-to-one to the TS children based on sex and age. If more than one TD participant met the matching criteria for a participant with TS, we selected the one who was closest to the participant with TS in age measured in months and were in the same school grade. The individuals in the pairs did not differ more than six months in age and were in the same school grade. None of the matched TD children had any psychiatric, neurological, or neurodevelopmental disorders according to parental reports. All participants were native Hungarian speakers, and they had normal or corrected-to-normal vision. Table 2.1. summarizes the descriptive characteristics of the groups alongside with other cognitive measurements often reported in TD (Robertson et al., 2017).

The experiment composed of two sessions on the same day with a 5-hour delay between them. In this study, we report only the first part of the experiment analyzing the learning phase of the procedural learning task. Parents of all participants were asked to complete the Strength and Difficulties Questionnaire (SDQ; Goodman, 1997) to measure hyperactivity, emotional difficulties, conduct, and peer problems. Caregivers of all participants provided written consent and children assented to participate in the study before testing. The study was approved by the local institutional research ethics committee and was conducted in accordance with the Declaration of Helsinki.

Table 2.1. Descriptive data of the participants and performance on the cognitive measurements.

	Group					
	TD (n = 21)		TS (n = 21)			
	M	*SD*	*M*	*SD*	*t*	*p*
Age in months	149.38	16.98	148.43	16.41	0.19	0.85
School grade	6.00	1.34	5.90	1.34	0.61[1]	0.96
SDQ total score	8.38	4.64	11.42	6.41	-1.73	0.09
YGTSS total score	–	–	17.43	8.12	–	–
WCST perseverative error (%)	16.42	8.64	16.02	7.53	0.16	0.87
Phonemic verbal fluency	10.76	3.03	10.17	2.74	0.67	0.51
Semantic verbal fluency	19.79	5.30	18.74	3.40	0.76	0.45
Counting span	3.38	0.60	3.67	0.97	-1.11	0.28

Note. The neuropsychological tests are well-known tasks for measuring executive functions. Wisconsin Card Sorting Task (Berg, 1948; Mueller & Piper, 2014) was used to measure cognitive flexibility. A higher percentage of perseverative errors indicate worse cognitive flexibility. Phonemic and semantic verbal fluency (Strauss, Sherman, & Spreen, 2006; for Hungarian version, see Tanczos, Janacsek, & Nemeth, 2013a; Tanczos, Janacsek, & Nemeth, 2013b) measures the central executive component of the working memory model. Here, verbal fluency is measured by the number of correct words. The counting span task (Case, Kurland, & Goldberg, 1982) is a complex working memory task. Participants' counting span capacity is calculated by the highest set size they were able to recall in the correct order. SDQ = Strength and Difficulties Questionnaire. YGTSS = Yale Global Tic Severity Scale. WCST = Wisconsin Card Sorting Task. [1]G-test was used instead of Chi-square test as the assumptions of Chi-square test were not met.

2.2. Tasks

2.2.1. Alternating Serial Reaction Time (ASRT) task

Statistical and sequence learning was measured by the cued version of the Alternating Serial Reaction Time (ASRT) task (Nemeth, Janacsek, & Fiser, 2013). In this task, a target stimulus (either a dog's head or a penguin) appeared in one of the four equally spaced, horizontally arranged possible locations (empty circles). Participants were instructed to press the corresponding key on the keyboard (Z, C, B or M) as accurately and as fast as they could. The stimulus remained on the screen until the participants responded, then, after a 120-ms-long delay, the next target appeared.

The presentation of the stimuli followed an 8-element alternating sequence where pattern and random elements alternated with each other (e.g., 1-r-2-r-4-r-3-r, where numbers represent the locations from left to right and 'r' indicates a randomly selected location). In the cued ASRT task, the pattern and random elements are visually distinguishable, pattern elements are indicated by the dog's head and random elements are indicated by the penguins. Participants were informed about the presence of the sequence structure, they were told that the dogs always follow a predetermined pattern, while penguins appear randomly in one of the possible locations. They were instructed to find the pattern of the dog's appearance to improve their performance. The alternation of pattern and random elements creates six unique sequence permutations: 1-r-2-r-3-r-4-r, 1-r-2-r-4-r-3-r, 1-r-3-r-2-r-4-r, 1-r-3-r-4-r-2-r, 1-r-4-r-2-r-3-r, and 1-r-4-r-3-r-2-r. Note that each of these six permutations can start at any location (e.g., 1-r-2-r-3-r-4-r and 2-r-3-r-4-r-1-r are identical sequence permutations). One of the permutations were selected to each participant and it was counterbalanced across participants in each group.

The structure of the ASRT task results in some runs of three successive elements – referred to as *triplets* – more frequent than others. If the sequence is 1-r-2-r-4-r-3-r, triplets such as 1-X-2, 2-X-4, 4-X-3, 3-X-1 (X indicates the middle element of the triplet) occur often since their last element can be either pattern or random. However, 3-X-2 or 4-X-2 occur less frequently as the third element could only be random. The more frequent triplet types are labeled as "high-frequency" triplets, while the latter types are labeled as "low-frequency" triplets (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013). The labels also refer to the transitional probabilities inside the triplets meaning that the third element of a high-frequency triplet is highly predictable from the first element of the triplet (with 62.5% probability). In case of the low-frequency triplet, the predictability of the final element is lower (12.5%).

Furthermore, each element can be categorized by their structure, meaning whether they are pattern or random elements (note that these are differentiated by visual cues). We can distinguish high-frequency triplets with the last element being a pattern element and high-frequency triplets with the last element being a random element. The last element of a low-frequency triplet can only be a random element as pattern elements always appear with high probability.

Previous studies have shown that participants perform differently on the different triplets. Participants show faster performance on the high-frequency triplets compared to the low-frequency ones (e.g., Howard, & Howard, 1997; Janacsek et al., 2012; Takács et al., 2018), and they also show faster performance on pattern triplets compared to the random ones (e.g., Howard, & Howard, 1997; Kóbor et al., 2018; Simor et al., 2019). Therefore, we can differentiate three trial types: (1) trials that belongs to the predetermined sequence and are the last element of a high-frequency triplet called *pattern* trials, (2) trials that appear randomly and also are the last element of a high-frequency triplet called *random high* trials, and (3) random elements that appear as the final element of a low-frequency triplet labeled as *random low* trials.

Different performance on these trial types can help differentiating the two aspects of procedural learning, sequence learning and statistical learning (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013). Sequence learning is measured by the difference in reaction times between pattern and random high elements. These elements share the same statistical properties as they both correspond to the last element of a high-frequency triplet, but they differ in sequence properties as one of them is part of the predefined sequence while the other appears randomly. Therefore, faster response to the pattern compared to the random high trials indicates greater sequence learning. To assess statistical learning, we compare the reaction times between random high and random low trials. Here, the elements share the sequence structure (both are random) but differ in statistical properties as they correspond either to the final element of a high-frequency or a low-frequency triplet. Therefore, greater statistical learning is defined as faster reaction time to random high than to random low elements. To sum up, statistical learning grasps purely frequency-based learning, whereas sequence learning shows the acquisition of order information. The structure of the ASRT task and the quantification of the underlying learning mechanisms illustrated in Fig. 2.1.

The task consisted of 20 1-minute-long blocks, each block contained 85 trials. Each block started with 5 random trials for practice, then the unique 8-element alternating

sequence was presented 10 times. After each block, awareness of the sequence structure was measured. Participants were instructed to type the order of the dog's head using the corresponding keys. The sequence report lasted until the participants produced 12 consecutive responses, ideally, the given 4-element sequence three times. This method allowed us to determine the duration (in terms of blocks) participants needed to discover the sequence as defined by consistently reporting the given sequence with an at least 70% accuracy from that point. We labeled this variable as the *timing of the discovery* of the sequence. We also quantified the average knowledge about the sequence formed by the end of the task. We used each reported sequence after the last five blocks and calculated how many responses out of the 12 was correct after each block. Then the mean of these percentile variables was calculated for each participant. We labeled this average of the reports as *explicit knowledge*.

A

Sequence structure: e.g., 1 – r – 2 – r – 4 – r – 3 – r

P	r	P	r	P	r	P	r	P	r	P	r	P	r	P
1	1 2 3 4	2	1 2 3 4	4	1 2 3 4	3	1 2 3 4	1	1 2 3 4	2	1 2 3 4	4	1 2 3 4	3

B

	Structure: P-r-P (the last element is always pattern)	Structure: r-P-r (the last element is always random)
High-frequency triplets (62.5% of all trials)	3-4-**1** (50%)	3-4-**1** (12.5%)
Low-frequency triplets (37.5% of all trials)	never occurring (always high)	3-4-**2** (12.5%) 3-4-**3** (12.5%) 3-4-**4** (12.5%)

Sequence learning: pattern high – random high

Statistical learning: random high – random low

Figure 2.1. An example of sequence structure, (A) triplet types and the underlying learning mechanisms (B) in the cued Alternating Serial Reaction Time (ASRT) task. In the example of the alternating sequence structure (A), numbers indicate pattern elements and 'r' indicates a randomly selected location. The alternating sequence makes some runs of three consecutive elements more frequent than others. Based on the structure, among high-frequency triplets, we can differentiate pattern high triplets (with red shading in Fig.2.1A and red font in Fig.2.1B) and random high triplets (with gold shading in Fig.2.1A and gold font in Fig.2.1B). Low-frequency triplets can only end with a random element (random low triplets, with blue shading in Fig.2.1A and blue font in Fig1B). Statistical learning is quantified by contrasting the reaction time of the random high and random low triplets (gold vs. blue, the right column of the table). Sequence learning is quantified by contrasting the reaction time of the pattern high and random high triplets (red vs. gold, the top row of the table). Adapted from Nemeth, Janacsek, and Fiser (2013).

2.2.2. Yale Global Tic Severity Scale (YGTSS)

Tic severity was measured by the Yale Global Tic Severity Scale (Leckman et al., 1989), which is a reliable and conventional measurement of tic severity. YGTSS is a semi-structured interview, which rates motor and vocal tics based on their number, frequency, complexity, intensity, and interference with everyday life on a scale of zero to five for motor and phonic tics individually. The Total Score reported here consists of the motor and phonic scores with a maximum score of 50. Tic severity was measured regarding symptoms in the last week.

2.3. Data analysis

Statistical analysis of the ASRT task was based on previous studies (Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019). The 20 1-minute-long blocks were collapsed into four epochs, each containing five blocks. Each trial was categorized as the final element of a *pattern high*, *random high* or *random low* triplet. Two types of low-frequency triplet were excluded from the analysis, repetitions (e.g., 222, 444) and trills (e.g., 121, 242), since participants often show pre-existing response tendencies to these items (Song et al., 2007a, 2007b). The median of RT data (for correct responses) was calculated for each participant in each epoch, separately for the three types of triplets. We also calculated learning scores separately for each epoch for the two types of underlying learning processes. Statistical learning scores were calculated as the difference in RT between random high and random low triplets, while sequence learning scores were calculated as the difference in RT between pattern high and random high triplets.

To examine the two learning mechanisms, RT data were analyzed in a mixed design ANOVA across the four epochs. Statistical learning was quantified with a mixed-design ANOVA with FREQUENCY (random high-frequency and random low-frequency triplets) and EPOCH (1-4) as within-subjects factors and GROUP (TS and TD) as a between-subjects factor. Sequence learning was also quantified with a mixed-design ANOVA with ORDER (pattern high-frequency and random high-frequency triplets) and EPOCH (1-4) as within-subjects factors and GROUP (TS and TD) as a between-subjects factor. To test for post hoc pairwise comparisons, we used LSD (Least Significance Difference) tests. The Greenhouse-Geisser epsilon correction was used when necessary. As a measure of effect size partial eta-squared (η^2_p) is reported.

3. Results

To compare **statistical learning** among the TS and TD groups, we conducted a mixed-design ANOVA on the RT (see Data analysis). The ANOVA revealed a significant FREQUENCY main effect ($F(1, 40) = 71.4$, $p < .001$, $\eta^2_p = 0.64$), meaning that RTs were faster on random high-frequency triplets compared with random low-frequency triplets. The main effect of EPOCH was also significant ($F(3,120) = 44.90$, $p < .001$, $\eta^2_p = 0.53$), indicating that, over groups, participants became faster with practice on both random high and random low-frequency triplets. The FREQUENCY*GROUP interaction was at the trend-level ($F(1, 40) = 3.31$, $p = .076$, $\eta^2_p = 0.076$), while FREQUENCY*EPOCH*GROUP interaction was significant ($F(3, 120) = 2.96$, $p = .035$, $\eta^2_p = 0.07$), indicating that the time course of statistical learning was different between the groups (see Fig. 2.2). Follow-up analysis on the statistical learning score revealed a difference in the first epoch between the groups: The TS group showed higher learning than the TD group (TS: $M = 27.38$ ms, $SD = 31.45$ ms; TD: $M = -0.79$ ms, $SD = 28.91$ ms; $p = .004$; see Fig. 2.2C). There was no difference in learning in the remaining epochs (all ps > .203). The main effect of GROUP and other interactions were not significant (all ps > .291). In order to investigate whether the inclusion of ADHD and OCD comorbidities in the TS group could confound the results, we conducted the same analysis as above on the 17 children with TS without any comorbidities and the matched TD group. The analysis showed the same results as described above, indicating that the inclusion of TS participants with ADHD and OCD comorbidities does not confound the results (see Supplementary Material).

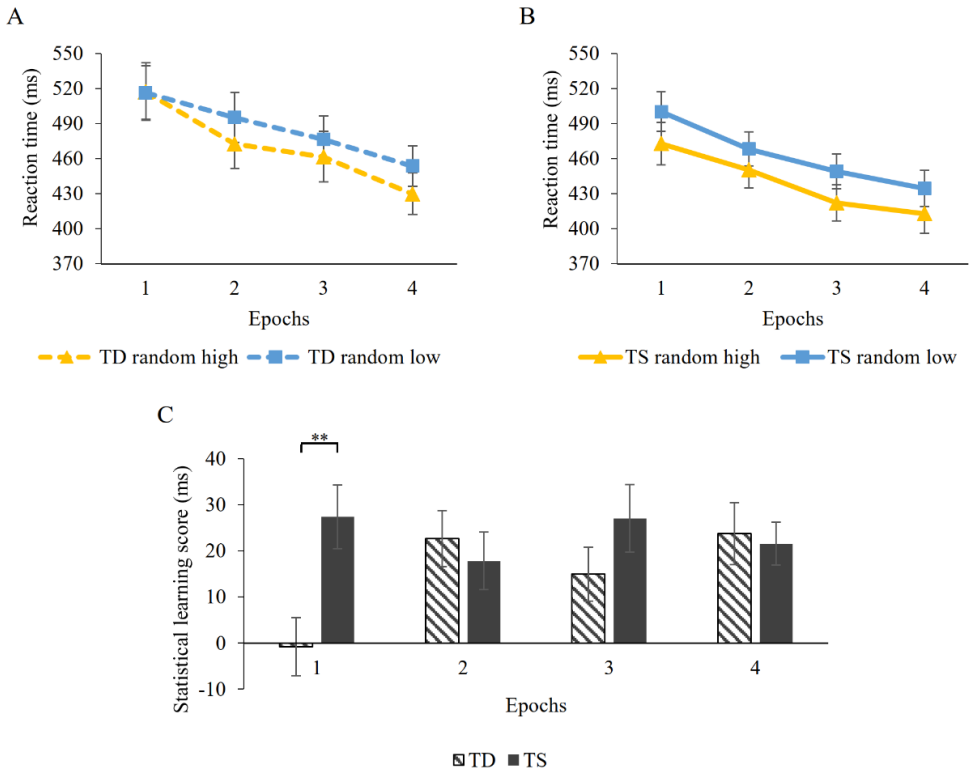

Figure 2.2. Statistical learning in the TD (A) and TS group (B). Dashed lines represent the TD group, continuous lines represent the TS group. Blue lines with square symbols indicate the reaction time (ms) on the random low triplets, gold lines with triangle symbols indicate the reaction time (ms) on the random high triplets. Statistical learning is indicated by the distance between the blue and gold lines. (C) Statistical learning score in the TD and TS group. Dashed bars represent the TD group, filled bars represent the TS group. Statistical learning score is computed by extracting reaction time of the random high triplets from reaction time of the random low triplets, separately in each epoch. Higher learning score indicates better learning. Error bars denote standard error of mean. $^{**}p < .01$.

To further examine the difference in statistical learning in the first epoch of the task between the TS and TD groups, we performed an additional analysis focusing on block-level data. We conducted a mixed design ANOVA with FREQUENCY (random high-frequency and random low-frequency triplets) and BLOCK (1-5) as within-subjects factors and GROUP (TS and TD) as a between-subjects factor. The main effect of FREQUENCY was significant ($F(1, 40) = 8.70$, $p = .005$, $\eta^2_p = 0.17$), participants showed faster RTs on the random high-frequency triplets compared to the random low-frequency triplets. The main effect of BLOCK was also significant ($F(4, 160) = 2.66$, $p = .035$, $\eta^2_p = 0.063$), suggesting reaction times became faster on both triplets with practice in both groups. The FREQUENCY*GROUP interaction was significant ($F(1, 40) = 14.24$, $p = .001$, $\eta^2_p = 0.263$), the TS group showed faster RTs to random high-frequency triplets than random low-frequency ones overall in the first epoch ($M = 32.30$, $SD = 32.86$), while

the TD group did not show learning in the first epoch (M = -3.96, SD = 29.31). Crucially, FREQUENCY*BLOCK*GROUP interaction was marginally significant (F(4, 160) = 2.17, p = .074, η^2_p = 0.05). The post hoc analysis revealed that the TD group showed similar RTs on random high-frequency and random low-frequency trials in four out of five blocks (ps > .206), and even marginally *faster* responses on random-low frequency compared to random-high frequency triplets in the remaining block (i.e., the opposite direction than expected for statistical learning; p = .073), suggesting that the TD group did not acquire the statistical knowledge in Epoch 1. In contrast, the TS group showed comparable RTs on both trial types only in the first block (p = .949), and showed marginally significant (Blocks 2 and 4, ps < .084) or significant (Blocks 3 and 5, ps < .026) statistical learning in the remaining blocks. This block-wise analysis provides evidence that the difference between the TD and TS groups in the epoch-wise analysis on statistical learning is not due to pre-existing response tendencies. Instead, it suggests that the TS group acquired the statistical knowledge gradually albeit early in the task (around Blocks 2-3), while the TD group required more practice to achieve a similar level of knowledge as the TS group (observed in the later epochs).

To investigate **sequence learning,** we also used a mixed design ANOVA on the RT (see Data analysis). The main effect of ORDER was significant (F(1, 40) = 8.35, p = .006, η^2_p = 0.17), suggesting that participants showed faster RTs on pattern high-frequency triplets compared with random high-frequency ones. The main effect of EPOCH was also significant (F(2.3, 90.1) = 42.33, p < .001, η^2_p = 0.51), indicating that participants became faster with practice on both triplets. The significant ORDER*GROUP interaction (F(1, 40) = 4.93, p = .032, η^2_p = 0.11) suggests that the two groups differed in the RT difference between the triplets. Follow-up analysis on the learning scores showed that the TD group learned to differentiate between pattern high and random high-frequency triplets, but the TS group showed similar RTs on both triplets (TD: M = 38.46 ms, SD = 66.11 ms; TS: M = 5.04 ms, SD = 19.71 ms) (see Fig. 2.3). The EPOCH*GROUP interaction was at the trend-level (F(2.3, 90.1) = 2.67, p = .068, η^2_p = 0.063), other main effects or interactions were not significant (all ps > .223). We conducted the same analysis as above on the TS group without any comorbitidies and the matched TD group to investigate whether comorbidities could confound the results. The analysis without comorbidities showed identical results as the analysis involving TS participants with ADHD and OCD comorbidities (see Supplementary Material).

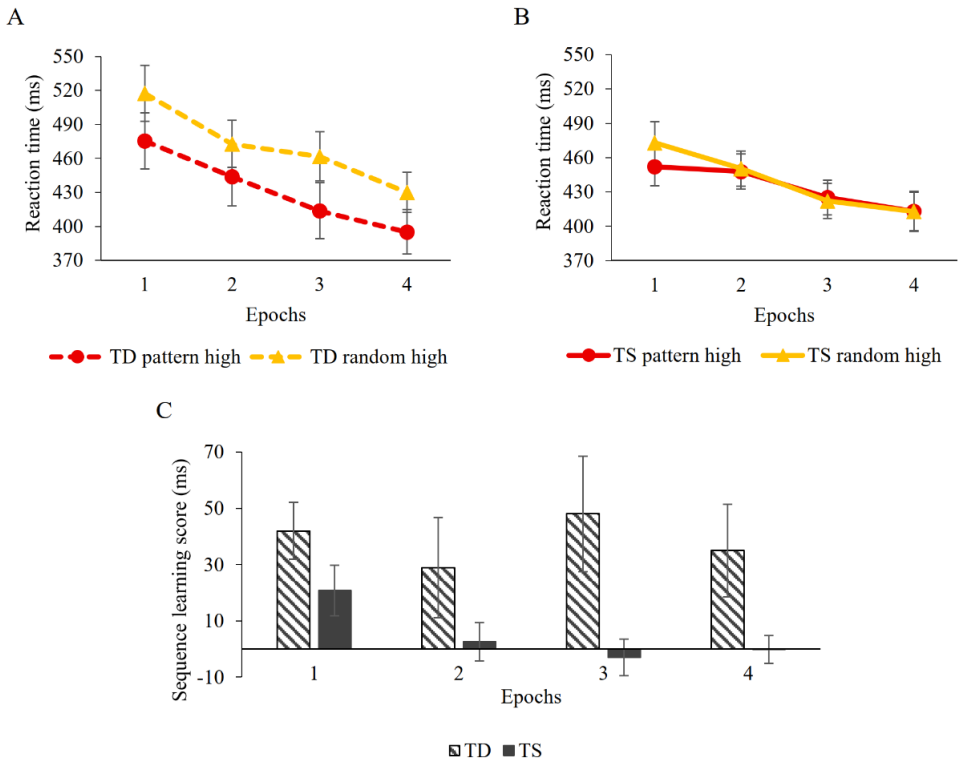

Figure 2.3. Sequence learning in the TD (A) and TS group (B). Dashed lines represent the TD group, continuous lines represent the TS group. Red lines with circle symbols indicate the reaction time (ms) on the pattern high triplets, gold lines with triangle symbols indicate the reaction time (ms) on the random high triplets. Sequence learning is indicated by the distance between the red and gold lines. (C) Sequence learning score in the TD and TS group. Dashed bars represent the TD group, filled bars represent the TS group. Sequence learning score is computed by extracting reaction time of the pattern high triplets from reaction time of the random high triplets, separately in each epoch. Higher learning score indicates better learning. Error bars denote standard error of mean.

To evaluate the relationship between tic severity and procedural learning in the TS group, we correlated statistical and sequence learning scores with the YGTSS total score. First, we investigated **the relation between statistical learning and tic severity**. One participant showed extremely high statistical learning score according to Tukey's (1977) criterion (more than 1.5 times the interquartile range) and was an outlier with regard to the relation of statistical learning and tic severity. We excluded this participant from the correlation analysis. The analysis revealed a negative relationship at the trend-level (r = -.43, p = .06), suggesting better statistical learning in children with less severe tics (Fig. 2.4A). The **correlation between sequence learning and tic severity** was not significant (r = .18, p = .44; Fig. 2.4B).

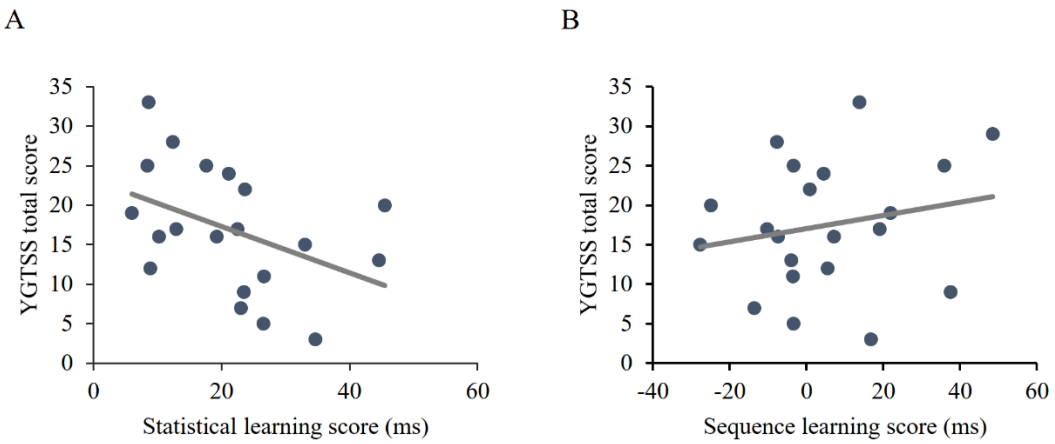

Figure 2.4. Correlation between (A) YGTSS total score and statistical learning score and (B) between YGTSS total score and sequence learning score. YGTSS = Yale Global Tic Severity Scale. Statistical learning score is the difference in RT between random high and random low-frequency triplets. Sequence learning score is the difference in RT between pattern high and random high-frequency triplets.

In order to check whether participants followed the instruction to find the predetermined 4-element sequence of the pattern stimulus' (dog's head) appearance, we asked them to report the sequence of the dog's head after each block. According to the results, explicit knowledge about the sequence was present early in the task, the timing of discovery was around the 6th block, and it did not differ between the TS and TD groups ($t(34) = 0.199$, $p = .843$; $M_{TS} = 5.61$, $SD_{TS} = 6.55$; $M_{TD} = 6.05$, $SD_{TD} = 6.82$). Explicit knowledge of the sequence also suggests that the participants followed the instructions, the mean explicit knowledge score in the last epoch was 89% (SD = 20%) in the TS and 79% (SD = 29%) in the TD group. Moreover, we found no significant difference between the groups ($t(40) = -1.25$; $p = .218$), suggesting that similar explicit knowledge emerged in the groups about the predetermined sequence structure.

4. Discussion

The goal of the present study was to examine how two aspects of procedural learning, namely statistical and sequence learning, are affected in TS, and test whether these aspects contribute to the procedural hyperfunctioning observed in previous studies. We used the cued version of the Alternating Serial Reaction Time task, which allowed us to examine the two aspects simultaneously, in the same experimental design. We found enhanced sensitivity to statistical information in TS, while the TS group showed impaired

sequence learning. Furthermore, executive functions and working memory capacity did not differ between the groups (Table 2.1).

4.1. Sensitivity to statistical information

Children with TS showed enhanced sensitivity to statistical information compared to their typically developing peers. This result is in line with previous studies showing speeded processing on tasks tapping into procedural learning and memory (Dye et al., 2016; Shephard et al., 2019; Takács et al., 2018; Walenski et al., 2007). In the present study, the enhanced sensitivity to statistical information was more prominent at the beginning of the task. The steepness of the learning curve is a sensitive index of how learning occurs in a specific group (Barnes, Howard, Howard, Kenealy, & Vaidya, 2010). Prior studies on neurotypical population showed that statistical learning reach a plateau early, hence, probabilistic information is learned rapidly and then remains stable (Kóbor et al., 2018; Simor et al., 2019). Our results showed the same pattern in both groups, however, it happened faster in the TS than in the TD group. Similar pattern was reported in the study of Takács et al. (2018), in which children with TS showed better learning at the end of the first learning session than their TD peers. The study of Takács et al. (2018) employed a probabilistic sequence learning task, in which participants acquire probabilistic information in an incidental manner. In that task version with non-cued stimuli, learning may occur in a slower pace than in the cued version of the task (note that the knowledge of statistical information remains consciously inaccessible to participants even in the cued version of the task; for more details, see Simor et al., 2019). Thus, faster procedural learning in TS can be found in both non-cued and cued learning situations, and in a more speeded manner in the latter case.

Our result is in line with previous studies that used tasks with probabilistic sequence structure (Shephard et al., 2019; Takács et al., 2018). While these tasks were linked to procedural learning processes, it was not clear whether sensitivity to sequential or to probabilistic information (statistical learning) led to the enhanced performance. According to the results of the present study, enhanced sensitivity to probabilistic information may contribute to procedural hyperfunctioning. This is in line with the notion of procedural hyperfunctioning in TS, supported by Takács et al. (2018) and Shephard et al. (2019). The probabilistic sequence learning measured by the study of Takács et al. (2018) and statistical learning investigated in the present study are highly similar, as both require the acquisition of frequency-based information. Moreover, in the study of

Shephard et al. (2019), children with TS showed difficulties with transitioning from sequence to non-sequence blocks in the SRT task, indicating hyperlearning.

Sensitivity to statistical information might also explain the results of studies showing enhanced performance on language-based tasks (Dye et al., 2016; Walenski et al., 2007). The acquisition of complex probabilistic regularities (extraction of 2nd order non-adjacent transitional probabilities) has been found to be crucial in language acquisition and processing (Conway, Bauernschmidt, Huang, & Pisoni, 2010; Misyak, Christiansen, & Tomblin, 2010; Nemeth et al., 2011; Saffran et al., 1996; Thompson & Newport, 2007). We can see this, for instance, in studies showing that transitional probabilities between pairs of syllables are essential in the detection of word boundaries (Saffran et al., 1996). Processing statistical information is also part of syntactic processing, as transitional probabilities between word-like units help to detect phrase boundaries (Thompson & Newport, 2007). Moreover, processing of non-adjacent dependencies and individual differences in statistical learning are associated with differences in language ability (Misyak et al., 2010). Nemeth et al. (2011) also found evidence for the relation of statistical learning (with non-adjacent dependencies) and sentence processing in healthy adults. Kidd (2012) also reported an empirical demonstration of the association between statistical learning and syntactic processing in children. These results suggest that the processing of statistical information is important in language acquisition and language processing from infancy to adulthood.

Therefore, the "speeded grammatical processing" seen in children with TS (Dye et al., 2016; Walenski et al., 2007) could reflect their enhanced sensitivity to statistical information. In detail, the non-word repetition task used in the study of Dye et al. (2016) involves rule-governed (de)composition of the non-words. Participants do not simply repeat the non-words, they separate them into phonological segments then attempt to reconstruct them. This process is influenced by the phonotactical constraints of the language (see e.g., Coady & Evans, 2008). Acquiring and using phonotactical constraints within words requires detecting and using transitional statistics, i.e., statistical learning. Moreover, the faster production of regular past tenses in TS (Walenski et al., 2007) can also reflect enhanced sensitivity to statistical information, as rule-governed composition of morphemes also involves transitional statistics. However, it is not clear how enhanced sensitivity to probabilistic information can explain the speeded tool but not animal naming in TS (Walenski et al., 2007). It might be possible that better statistical learning

has an additive effect, which can be generalized to tool naming (and perhaps to other procedural functions), but further studies are warranted to investigate this notion.

One can argue that other alternative explanations may better explain the group differences found in the present study. One such alternative explanation could be a general "speeded processing" in TS, that could be captured by generally faster reaction times, irrespective of the stimulus types. However, generally faster reaction times cannot alone explain the accumulating evidence of procedural hyperfunctioning in TS. Takács et al. (2018) found that children with TS showed more prediction errors indicating enhanced procedural functions, while in the study of Shephard et al. (2019) children with TS showed difficulties with the transition from sequenced to non-sequenced learning. None of these learning measures are directly related to "speeded processing". Furthermore, general "speeded processing" cannot explain the findings of Walenski et al. (2007). If children with TS had overall faster response times, it would have manifested not only in producing regular past tenses and tool naming (related to procedural functions) but also in producing irregular past tenses and animal naming (related to declarative functions) (Walenski et al., 2007). However, the study of Walenski et al. (2007) showed no differences between the TS and TD groups in response times related to declarative functions, suggesting that general "speeded processing" cannot explain their results. Moreover, our current results also do not support generally faster processing or response execution as we did not find a difference between the TS and the control group in average reaction times (indicated by the non-significant main effect of Group; see Results). Note that the general "speeded processing" notion detailed above is different from the Clinical Extension Hypothesis, introduced in the study of Dye et al. (2016). This hypothesis proposes that neurobiological alterations in TS might result in speeded performance supported by processes related to the neurobiological alterations.

4.2. Sensitivity to sequential information

Our results indicate impaired sequence learning in TS as children did not differentiate between pattern high-frequency and random high-frequency triplets. Similar alterations have been demonstrated in previous studies of motor learning in TS (Avanzino et al., 2011; Palminteri et al., 2011; Stebbins et al., 1995). Avanzino et al. (2011) found impaired performance in TS children on a sequential single-hand finger-tapping task. Similarly, Stebbins et al. (1995) reported deficits in a motor learning task in TS. Furthermore, in the study of Palminteri et al. (2011), TS and TD showed different

performance in a motor skill learning task, where triplets were associated with different outcomes: high reward or minimal reward. While participants with TS showed enhanced learning in the high reward condition, the group difference was reversed when only minimal reward was present. Therefore, possible sequence learning impairment can be modified or even masked by other involved processes such as the reward system. In our study, the task was designed to clearly differentiate between sequence learning and sensitivity to probabilistic information.

Accumulating evidence based on neurotypical population suggests a dissociation between statistical learning (processing of frequency-based information) and sequence learning (processing of serial order information). First, previous studies suggested at least partially different developmental trajectories of statistical and sequence learning (Nemeth, Janacsek, & Fiser, 2013). Second, while statistical information is typically acquired relatively rapidly and incidentally, sequence learning seems to occur with gradually improving performance, irrespective of whether it occurs incidentally or intentionally (at least, when measured with the ASRT task; Howard, Howard, Japikse, DiYanni, et al., 2004; Howard, & Howard, 1997; Simor et al., 2019). Furthermore, they are distinguishable on the neural level as suggested by event-related potentials during learning (Kóbor et al., 2018) and by neural oscillations during consolidation (Simor et al., 2019). The present study further supports the notion of multifactorial procedural learning, as we found a dissociation between two aspects of learning in the clinical group: the TS group showed enhanced statistical learning and impaired sequence learning on the ASRT task. It is possible that the two aspects of learning compete with one another in TS. Therefore, having enhanced processing on one of them results in having a disadvantage on the other. Future studies are warranted to test this possibility.

Note that we used a cued version of the ASRT task since previous studies showed relatively faster acquisition of the alternating sequence in this task version (e.g., Kóbor et al., 2018; Simor et al., 2019), enabling to measure statistical and sequence learning in the same time frame (i.e., within one learning session). Consequently, while statistical learning occurred incidentally in the current study, sequence learning could have been supported by incidental as well as intentional learning processes. The intention to learn may have interfered with the acquisition of the alternating sequence selectively in the TS group but not in the TD group. This interpretation, however, seems unlikely. Both groups showed similar working memory and executive function capacity (see Table 2.1), suggesting similar cognitive resources that are required to follow the instructions in the

task. Indeed, both groups acquired similar level of explicit knowledge about the sequence (as measured by the *sequence reports* after each block). The weaker performance in the TS group appeared to be limited to weaker sequence learning as measured by the reaction time *learning scores*. Additionally, the fast pace of the task (typical responses under 500 ms) makes it difficult for the consciously accessible sequence knowledge to substantially influence participants' response times, leading to at least somewhat dissociable measures (Horvath et al., 2020). In this view, the sequence report may serve as a more explicit measure of sequence knowledge, and the reaction time learning scores may reflect a more implicit, incidental measure of sequence knowledge, even in an intentional learning situation. This pattern of findings suggests that, even though participants performed a cued version of the task and had intention to learn the alternating sequence, the TS group's weaker performance may be selective to the implicit measure of sequence acquisition, irrespective of whether learning occurs incidentally or intentionally. Nevertheless, future studies are needed to directly test this possibility.

4.3. Procedural functioning and symptom severity in TS

We tested the spectrum of tic severity using the Yale Global Tic Severity Scale. Sensitivity to statistical information showed marginally significant negative correlation with the severity of tics, indicating that enhanced sensitivity to statistical information can emerge in conjunction with less severe tics. Besides tic severity, premonitory urges could also relate to sensitivity to statistical information. Premonitory urges are described as a feeling of tightness or tension resulting in discomfort or distress and only can be relieved by performing specific tics (Robertson et al., 2017). However, the relation between premonitory urges and tics is not deterministic, tics can be present without premonitory urges. It has been shown that premonitory urges are associated with interoceptive awareness (Ganos et al., 2015). Interoceptive information is processed implicitly. Being highly sensitive to implicit statistical information could lead to being more aware of or sensitive to premonitory urges. The relation between procedural hyperfunctioning and sensitivity to premonitory urges may also converge on the neural level as supplemental motor area is important in both processes (Conceição et al., 2017; Grafton, Hazeltine, & Ivry, 2002; Leckman & Cohen, 1999; Peterson, 1999). Future studies should explore this connection between sensitivity to statistical information and to premonitory urges, especially considering the importance of premonitory urge detection in therapy (see habit reversal training, Piacentini & Chang, 2005).

4.4. Limitations and clinical implications

The finding of the present study is limited to a specific TS population, namely, those with less severe symptoms and without comorbidities. In our study, participants with TS are characterized with mild to moderate symptoms, indicated by the YGTSS. Future studies should test whether procedural hyperfunctioning is present in children with severe symptoms. Additionally, most of the children in the clinical group had TS without comorbidities (only 3 children had comorbid ADHD and 1 comorbid ADHD and OCD), therefore, sensitivity to statistical information seems to be specific to TS. Comorbidities can contribute to a greater interindividual variability in procedural functions and may mask the differences specifically related to TS. Future studies are warranted to examine sensitivity to statistical information in subgroups of TS population, such as TS with specific comorbidities. Future investigations also seem to be warranted on whether these findings extend to disorders with similar neurocognitive profiles as TS, such as OCD (Roth, Baribeau, Milovan, O'Connor, & Todorov, 2004).

Our study has both clinical and educational implications. Procedural memory plays an important role in skill acquisition, such as sports, language, or even social skills (Frith & Frith, 2012; Kaufman et al., 2010; Lieberman, 2000). Strong skill-based competencies in TS could help reduce the disadvantages related to the disorder. Moreover, skill-based training using frequency-based information might also help in reducing behavioral symptoms and learning disadvantages. Future studies are needed to develop such training methods or improve already existing ones, and to test their effects in practice.

4.5. Conclusion

In the present study, our aim was to investigate two aspects of procedural learning, namely statistical and sequence learning, and test whether these aspects of learning contribute to the procedural hyperfunctioning in TS proposed by previous studies. Our results showed further evidence for enhanced procedural functions in TS with a heightened sensitivity to statistical information, while sequence learning was impaired in TS. These results suggest that sensitivity to frequency-based information may contribute to the procedural hyperfunctioning in TS, shedding light on a cognitive advantage in TS.

III. Study 3: Access to procedural memories after one year: evidence for robust memory consolidation in Tourette syndrome

Publication:

Tóth-Fáber, E., Tárnok, Z., Takács, Á., Janacsek, K., & Nemeth, D. (2021). Access to procedural memories after one year: evidence for robust memory consolidation in Tourette syndrome. *Frontiers in Human Neuroscience*, 458.
https://doi.org/10.3389/fnhum.2021.715254

Abstract

Tourette syndrome is a childhood-onset neurodevelopmental disorder characterized by motor and vocal tics. On the neural level, tics are thought to be related to the disturbances of the cortico-basal ganglia-thalamo-cortical loops, which also play an important role in procedural learning. Several studies have investigated the acquisition of procedural information and the access to established procedural information in TS. Based on these, the notion of procedural hyperfunctioning, i.e., enhanced procedural learning, has been proposed. However, one neglected area is the retention of acquired procedural information, especially following a long-term offline period. Here, we investigated the five-hour and one-year consolidation of two aspects of procedural memory, namely serial-order and probability-based information. Nineteen children with TS between the ages of 10 and 15 as well as 19 typically developing gender- and age-matched controls were tested on a visuomotor four-choice reaction time task that enables the simultaneous assessment of the two aspects. They were retested on the same task five hours and one year later without any practice in the offline periods. Both groups successfully acquired and retained the probability-based information both when tested five hours and then one year later, with comparable performance between the TS and control groups. Children with TS did not acquire the serial-order information during the learning phase; hence, retention could not be reliably tested. Our study showed evidence for short-term and long-term retention of one aspect of procedural memory, namely probability-based information in TS, whereas learning of serial-order information might be impaired in this disorder.

1. Introduction

Tourette syndrome (TS) or Tourette Disorder is a childhood-onset neurodevelopmental disorder characterized by at least one vocal tic and multiple motor tics, which are not explained by medications or other medical conditions (APA, 2013). Tics can be expressed as simple or complex movements or vocalizations that are usually fast, abrupt, and semi-voluntary (APA, 2013). On the neural level, tics are thought to be related to the disturbances of the cortico-basal ganglia-thalamo-cortical (CBGTC) circuits (Albin et al., 2003; Albin & Mink, 2006; Maia & Frank, 2011; Mink, 2001; Peterson et al., 1998; Peterson et al., 2003; Stern et al., 2000). On the cognitive level, these circuits are also related to procedural learning (Doyon et al., 2009; Janacsek et al., 2020; Poldrack & Packard, 2003), which is considered to be the basis of skills and habits (Kaufman et al., 2010; Ullman, 2004). It has been proposed that tics and habits have similarities: both are stereotyped actions that are automatically executed and hard to inhibit (Conceição et al., 2017). Several studies have shown enhanced procedural learning, termed procedural hyperfunctioning, in TS (Dye et al., 2016; Shephard et al., 2019; Takács et al., 2018; Tóth-Fáber, Tárnok, et al., 2021). An important question emerges: does procedural hyperfunctioning in TS lead to persistent changes? Processing information does not stop at the end of a learning session, and long-term memory performance is based on the stabilization of encoded information, that is, on the consolidation of information (McGaugh, 2000; Walker, 2005). However, little is known about whether procedural hyperfunctioning persists over the consolidation periods and whether consolidation of procedural information differs in TS and neurotypical controls. In the present study, we focused on this question and investigated the short-term (five-hour) and long-term (one-year) consolidation of procedural information in children with TS.

A potential link has been suggested (Goodman et al., 2014; Takacs, Münchau, Nemeth, Roessner, & Beste, 2021) between procedural memory formation and habitual behavior both in everyday life and as a clinical phenomenon. Namely, tics that consist of sequential actions might rely on procedural memory associations. As mentioned above, similar neural networks are involved in the pathophysiology of TS and procedural learning. CBGTC circuits play a key role in the development of tics (Goodman et al., 2014; Worbe et al., 2010) and tics may result from a heightened direct pathway activity relative to the indirect pathway activity in the CBGTC loop (Maia & Frank, 2011; Mink, 2001). Tic-related activation has been shown in the premotor cortex and sensorimotor

cortex (Bohlhalter et al., 2006; Stern et al., 2000; Wang et al., 2011), in the supplementary motor area (Bohlhalter et al., 2006; Wang et al., 2011; Worbe et al., 2015), putamen (Bohlhalter et al., 2006; Stern et al., 2000; Wang et al., 2011), globus pallidus (Bohlhalter et al., 2006; Wang et al., 2011) and thalamus (Bohlhalter et al., 2006; Wang et al., 2011; Worbe et al., 2015). Accumulated evidence shows that these brain areas also play a role in procedural learning. Specifically, the formation of skills and habits in procedural memory has been linked to the basal ganglia, particularly to the striatum, and relies on the CBGTC loops (Doyon et al., 2009; Janacsek et al., 2020; Poldrack & Packard, 2003). Given the involvement of similar neural networks in procedural memory and the pathophysiology of TS, alterations of procedural functions can be expected in TS.

Procedural learning enables us to extract the regularities from the environment and underlies the acquisition and storage of skills and habits (Kaufman et al., 2010; Ullman, 2004). Humans are highly proficient in the extraction of transitional probabilities, that is, in the learning of predictive relations between events (i.e., the probability of event B following event A), even when these are nonadjacent (e.g., A – x – B, where the intervening event has no predictive value) (Conway, 2020; Frost & Monaghan, 2016). From a plethora in the environment, different kinds of regularities can be extracted. Two previously proposed regularities in relation to procedural memory are (1) serial order-based information and (2) probability-based, statistical information (Howard, Howard, Japikse, DiYanni, et al., 2004; Nemeth, Janacsek, & Fiser, 2013). Serial order-based information means that transitional probabilities between the elements are 1.0, which creates a deterministic serial order of events: for instance, event A is always followed by event B. Probability-based information refers to regularities where transitional probabilities are less than 1.0; here, higher transitional probability means higher predictability. Hence, extracting probability-based information enables the differentiation between more and less probable outcomes to learn stochastic relations between events: for instance, when event A is followed by event B in 75% of the cases and followed by event C in 25% of the cases. Although both regularities can be considered as learning of transitional probabilities (also often referred to as statistical learning), prior studies have shown considerable differences between them in healthy young adults (Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019; Takács et al., 2021). They revealed that the learning of serial-order regularities develops rather gradually, whereas the learning of probability-based regularities reaches its plateau in a quick manner (Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019),

which is also reflected in the neurophysiological correlates (Kóbor et al., 2018; Takács et al., 2021). In other words, the learning of serial-order regularities occurs relatively slowly, whereas participants acquire probability-based regularities rapidly and then show consistent, stable performance. Prior studies have also shown that successful acquisition of serial-order and probability-based information leads to the formation of long-term memory representations (Kóbor et al., 2017; Romano et al., 2010; Simor et al., 2019; Tóth-Fáber, Janacsek, & Németh, 2021; Zavecz et al., 2020). Thus, learning of these regularities might influence behavior on a longer timescale and outside of a lab environment as well.

Long-term memory performance is based on consolidation, that is, the stabilization of encoded memory representations (McGaugh, 2000; Walker, 2005). Empirically, consolidation is assessed by the difference in memory performance at the end of a session and at the beginning of the next one, following a delay (i.e., offline period). Consolidation can be revealed by successfully retained knowledge or by delayed gains of performance (i.e., offline learning) after the offline period (Robertson et al., 2004). Consolidation of any information is a complex process, which can be influenced by the encoded information, time (ultra-fast, short-or long-term consolidation), and the nature of the offline period (i.e., sleep or time spent awake) (Robertson, Pascual-Leone, & Press, 2004; Song et al., 2007b). Consolidation of serial-order and probability-based regularities has been investigated before, both over short-term (Simor et al., 2019; Zavecz et al., 2020) and long-term offline periods (Kóbor et al., 2017; Romano et al., 2010; Tóth-Fáber, Janacsek, et al., 2021). Romano et al. (2010) and Kóbor et al. (2017) focused on probability-based regularities in neurotypical adults, in both cases after a one-year offline period. Romano et al. (2010) showed successful retention of probability-based regularities in perceptual-motor skill experts (i.e., videogame and piano players) and non-experts. Kóbor et al. (2017) went beyond the study of Romano et al. (2010) by incorporating interference manipulation into their study design. They have demonstrated that memory representations of probability-based regularities are not only resistant to forgetting over a one-year offline period but are also resistant to interference. Furthermore, learning of serial-order and probability-based regularities seems to result in long-term memories in the developing mind, as well: Tóth-Fáber, Janacsek, et al. (2021) found evidence for one-year retention of such regularities in typically developing children and adolescents, thus extended the prior results on adults (Kóbor et al., 2017; Romano et al., 2010) to an age that is crucial in the development of procedural memory (Janacsek et

al., 2012; Juhasz et al., 2019; Zwart et al., 2019). In sum, memory representation of these regularities seems to be persistent over a long period of time both in neurotypical adults (Kóbor et al., 2017; Romano et al., 2010) and typically developing children (Tóth-Fáber, Janacsek, et al., 2021). Thus, it is possible to compare procedural memories between typical and atypical development after a long offline period. Crucially, extending the testing time to one year allows us to close the bridge between the time scale of lab experiments (typically hours or days) and real-world observations (i.e., when learning a new skill or developing a habit).

Several studies focused on procedural learning in TS with most showing intact (Channon et al., 2003; Takács et al., 2017) or even enhanced procedural functions (Dye et al., 2016; Shephard et al., 2019; Takács et al., 2018; Tóth-Fáber, Tárnok, et al., 2021; Walenski et al., 2007). Shephard et al. (2019), Takács et al. (2018) and Tóth-Fáber, Tárnok, et al. (2021) all employed variations of a well-known procedural learning task, the serial reaction time task (SRTT). Consequently, Shephard et al. (2019) showed enhanced learning of deterministic serial-order information, whereas Takács et al. (2018) and Tóth-Fáber, Tárnok, et al. (2021) showed enhanced learning of probability-based information in children and adolescents with TS. In conjunction with the online learning tasks, procedural hyperfunctioning has also been shown in tasks that measure the access to previously established procedural information, such as grammatical rules or vocabulary (Ullman, 2004; Walenski et al., 2007). For example, in the study of Walenski et al. (2007), compared to typically developing controls, children with TS showed faster production of rule-governed past tenses and faster naming of manipulated objects—both of which have been linked to procedural memory (Ullman, 2004). Additionally, Dye et al. (2016) provided evidence for enhanced access to established information in the phonological domain of language. They used a non-word repetition task that involved rule-governed grammatical (de)composition of non-words; therefore, it relied, at least in part, on procedural memory. Children with TS showed faster repetition of non-words than typically developing controls on this task. Two earlier studies reported findings on such tasks and showed faster access to established procedural information in TS in the morphological (Walenski et al., 2007) and phonological (Dye et al., 2016) domains of language. These findings suggest that not only procedural learning but also access to previously consolidated procedural knowledge may be enhanced in TS. This raises the question of whether the consolidation of procedural information is also atypical in this disorder.

Consolidation of procedural information in TS has not received much attention in previous research. Takács et al. (2018) incorporated a 16-hour offline period in their study design and investigated the learning and consolidation of probability-based regularities in children with TS. The TS group showed superior learning, but they showed greater forgetting following the overnight offline period than the typically developing controls. When controlling for the learning differences and comparing overnight changes as a function of prior knowledge, the TS and control groups showed comparable performance. Nevertheless, the differences in learning between the TS and typically developing groups make the interpretation difficult, and, as Takács et al. (2018) suggested, these results should be handled as inconclusive. According to our knowledge, there are no other studies up to date that directly investigated procedural consolidation in TS. In sum, it remains unresolved whether atypical procedural learning in TS leads to altered consolidation of procedural memories.

The present study focuses on the short- (five-hour) and long-term (one-year) consolidation of two aspects of procedural memory, namely serial-order and probability-based information in children with TS. To test this, we employed a widely used procedural learning task, namely the cued version of the Alternating Serial Reaction Time (ASRT) task, which enables us to measure the acquisition and consolidation of the two regularities simultaneously (Nemeth, Janacsek, & Fiser, 2013). Children with TS and age- and gender-matched typically developing controls performed the cued ASRT task in three sessions. To investigate the short-term consolidation of serial-order and probability-based information, the first two sessions took place on the same day with a 5-hour offline period between them. To test the one-year consolidation of the two regularities, the third session was administered following a one-year offline period. Hence, this explorative study aims to examine both the short-term and the long-term consolidation processes in children with TS.

2. Material and methods

2.1. Participants

Twenty children diagnosed with TS between the ages of 10 and 15 participated in our study. They were recruited through a child and adolescent psychiatry hospital in Budapest, Hungary. They had been diagnosed with TS based on the DSM-V criteria (APA, 2013). Diagnoses were made by a team of child psychiatrist, clinical psychologist

and special education teacher after a one-week-long observation in the hospital. One participant had to be excluded from the analyses as they consistently showed extremely low average accuracy on the regularity extraction task (more than 3 times the interquartile range from the quartiles; Tukey, 1977). Therefore, the final TS sample consisted of 19 children (16 boys and three girls). Demographic and clinical data of the TS participants are reported in Table 3.1. Three children had comorbid attention deficit hyperactivity disorder (ADHD) and one child had comorbid ADHD and obsessive-compulsive disorder (OCD). We did not exclude these participants from the analyses as ADHD and OCD are highly common in TS (Robertson et al., 2017). Participants did not have any other psychiatric or neurodevelopmental disorders. Three children were taking medication during either time of testing: one child was taking atomoxetine during the first testing, and two children were taking atomoxetine during the second testing. A subgroup of the TS children had been examined in the study of Tóth-Fáber, Tárnok, et al. (2021) (the overlap between the two samples is 81%), however, a new control group had been recruited due to difficulties in assessing the original control group one year later.

Seventy-eight typically developing (TD) children were recruited from local schools (note that the analyses on this sample had been reported in Tóth-Fáber, Janacsek, et al., 2021). From this group, we matched 19 children one-to-one to the TS participants based on age and gender (16 boys and three girls). The pairs had an age gap maximum of six months and were in the same school grade. None of the matched controls had any psychiatric, neurological, or neurodevelopmental disorders based on parental reports. All participants had normal or corrected-to-normal vision. Demographic data of the TD participants are reported in Table 3.1.

Caregivers of all participants completed a parental questionnaire regarding socioeconomic status (SES). SES was determined by the number of years the caregivers spent in formal education and it is reported in Table 3.1. Caregivers' average formal education was calculated based on both parents' education. In case of one participant in the TS group and three participants in the TD group, we only had information about one caregiver. In the TS group, data of two participants are missing.

Caregivers of all participants provided informed written consent, and children assented to participate in the study before enrollment. The study was approved by the research ethics committee of Eötvös Loránd University, Budapest, Hungary, and was conducted in accordance with the Declaration of Helsinki.

Table 3.1. Demographic and clinical data of the participants.

	Group			
	TS (*n* = 19)		TD (*n* = 19)	
	M	*SD*	*M*	*SD*
Age on the first testing day	11.95 years	1.27 years	11.79 years	1.48 years
School grade on the first testing day	5.68	1.29	5.95	1.47
Caregivers' average formal education	16.24 years	2.85 years	16.45 years	3.08 years
YGTSS total score on the first testing day	18.21	8.61	–	–
YGTSS total score on the second testing day	17.58	9.35	–	–

Note. YGTSS = Yale Global Tic Severity Scale.

2.2. Tasks

2.2.1. Alternating Serial Reaction Time (ASRT) task

The cued version of the Alternating Serial Reaction Time (ASRT) task (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013) was employed to measure the extraction of probability-based and serial-order regularities. The ASRT task has adequate test-retest reliability on neurotypical adult population (Stark-Inbar, Raza, Taylor, & Ivry, 2016). In this task, participants see four equally spaced empty circles which are horizontally arranged. A stimulus (either a dog's head or a penguin) occurs in one of the empty circles (Fig. 3.1A). Participants were instructed to press the corresponding key (Z, C, B, or M) on a QWERTY keyboard as accurately and as fast as they could. The response-to-stimulus interval was set to 120 ms.

In the task, pattern and random stimuli appeared in an alternating fashion. Pattern stimuli appeared following a predetermined sequence, whereas random stimuli could appear in one of the possible locations (i.e., empty circles). The stimuli were presented in blocks with 85 trials in each block. A block started with five random trials for practice, followed by an eight-element alternating sequence presented ten times. The alternating sequence consisted of pattern and random trials (e.g., 1-r-2-r-4-r-3-r, where numbers

indicate one of the four circles on the screen and ‘r’ indicates a randomly selected circle out of the four possible ones). In the cued ASRT task, participants are informed about the presence of the sequence, and their attention is drawn to the alternating sequence by marking the pattern and random trials with different visual stimuli. In our study, a picture of a dog denotes pattern trials, and a picture of a penguin represents random trials. Participants were not informed about the exact sequence, but they were instructed to find the pattern defined by the dogs’ appearance to improve their performance. For each participant, one of the six different sequence permutations was selected in a pseudo-random fashion, and the presence of the permutations was counterbalanced across participants and groups. For a given participant, the sequence permutation was the same across the epochs and the sessions. Note that the permutations can start at any location (e.g., 1-r-2-r-3-r-4-r and 2-r-3-r-4-r-1-r are identical sequence permutations).

Due to the alternating sequence (i.e., pattern and random elements occurring in an alternating fashion), some runs of three consecutive trials (referred to as triplets) were more probable than others. For example, if the sequence is 1-r-2-r-4-r-3-r, triplets such as 1-X-2, 2-X-4, 4-X-3, 3-X-1 (where X represents the middle element of the triplet) occur with a higher probability as their first and third elements could have been either pattern or random. This means that for example 4-X-3 can appear both as 4-2-3 (pattern – random – pattern) where the first and last elements are part of the predetermined sequence and as 4-2-3 (random – pattern – random) where the first and last elements are random, and the middle element is part of the predetermined sequence. However, triplets such as 3-X-2 or 4-X-2 were less probable as their first and third elements could have been only random (that is, random – pattern – random structure). More probable triplet types are referred to as “high-probability” triplets, while the less probable ones are labeled as “low-probability” triplets (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013). Each trial was categorized as the last element of either a high-probability or a low-probability triplet in a sliding window manner (i.e., one trial was the last element of a triplet, but it was also the middle and the first element of the two consecutive triplets, respectively). Another crucial feature of the trials is whether they belong to pattern or random elements (i.e., marked by a picture of a dog or a penguin). There are 64 unique triplets in the task, including all pattern-ending (50%) and random ending (50%) triplets. Sixteen of these unique triplets are high-probability and 48 triplets are low-probability. As high-probability triplets can occur as pattern-ending triplets (50% of all trials) and by ¼ chance as random-ending triplets (12.5% of all trials), these triplets constitute 62.5% of all trials

(Fig. 3.1C). Low-probability triplets constitute 37.5% of all trials. On the level of unique triplets, high-probability triplets are five times more probable than low-probability triplets (4% [62.5% / 16] vs. 0.8% [37.5% / 48]). In sum, three trial types could be differentiated: (1) trials that belonged to the predefined sequence and appeared as the last element of a high-probability triplet called *pattern trials*; (2) random elements that belong to a high-probability triplet labeled as *random high trials*; and (3) random elements that are the last element of a low-probability triplet called *random low trials* (Fig. 3.1B and 3.1C).

In the cued ASRT task, the acquisition of probability-based and serial-order regularities, also referred to as statistical and sequence learning, respectively, can be measured simultaneously (Nemeth, Janacsek, & Fiser, 2013). Learning of probability-based regularities was measured by the difference in reaction times (RTs) between random high and random low trials, greater learning was defined by faster RTs on random high than on random low trials. As both trials were random, they shared the same sequence properties but differed in statistical properties as they corresponded to the last element of a high-probability or a low-probability triplet, respectively. Learning of serial-order regularities was quantified by the difference in RTs between pattern and random high trials, greater learning was determined by faster RTs on pattern than on random high trials. The two trial types shared the statistical properties as they both were the last element of a high-probability triplet, however, they differed in sequence properties as pattern trials belonged to the predefined sequence (Fig. 3.1C).

At the beginning of the ASRT task, participants were instructed to discover the pattern of the dogs' appearance. At the end of each block, awareness of the serial-order structure was assessed. Participants were asked to type the order of the dogs using the corresponding keys. The post-block sequence report lasted until 12 consecutive responses, which ideally was the 4-element sequence three times. The post-block sequence reports after the last five blocks of the Learning Phase (see Procedure) were used to measure awareness of the sequence. We calculated how many out of the 12 consecutive responses were correct after each block; hence, we created a percentile variable. The mean of these five percentile variables was calculated for each participant, and we termed this variable as *explicit knowledge* of the sequence structure.

2.2.2. Yale Global Tic Severity Scale (YGTSS)

Tic severity was assessed by a widely used and conventional measurement, namely the Yale Global Tic Severity Scale (Leckman et al., 1989). YGTSS is a semi-

structured interview, which rates the number, frequency, complexity, intensity, and interference of motor and vocal tics separately on a scale of zero to five. The Total Score reported here contains the motor and vocal tic scores with a maximum score of 50. Tic severity was assessed two times: on the first testing day and one year later. Tic severity scores represent values from the week prior to the experiment.

2.3. Procedure

The study consisted of three sessions (Fig. 3.1D). The first two sessions took place on the same day with a 5-hour-long offline period between them. Children completed the learning session at the beginning of a school day and returned after their lunch break (that is, five hours later). The third session was administered ca. one year later (M_{delay} = 53.78 weeks, SD_{delay} = 3.11 weeks, between 47.95 and 60.57 weeks). Participants were assessed on the ASRT task in all three sessions. The ASRT task was presented in blocks. During the statistical analyses, blocks were collapsed into epochs, with each epoch containing five blocks. The Learning Phase consisted of 20 blocks (i.e., four epochs), the Testing Phase was composed of 10 blocks (i.e., two epochs) and the Retesting Phase again contained 20 blocks (i.e., four epochs). After the first testing day, participants were not informed that the ASRT task would be administered again one year later.

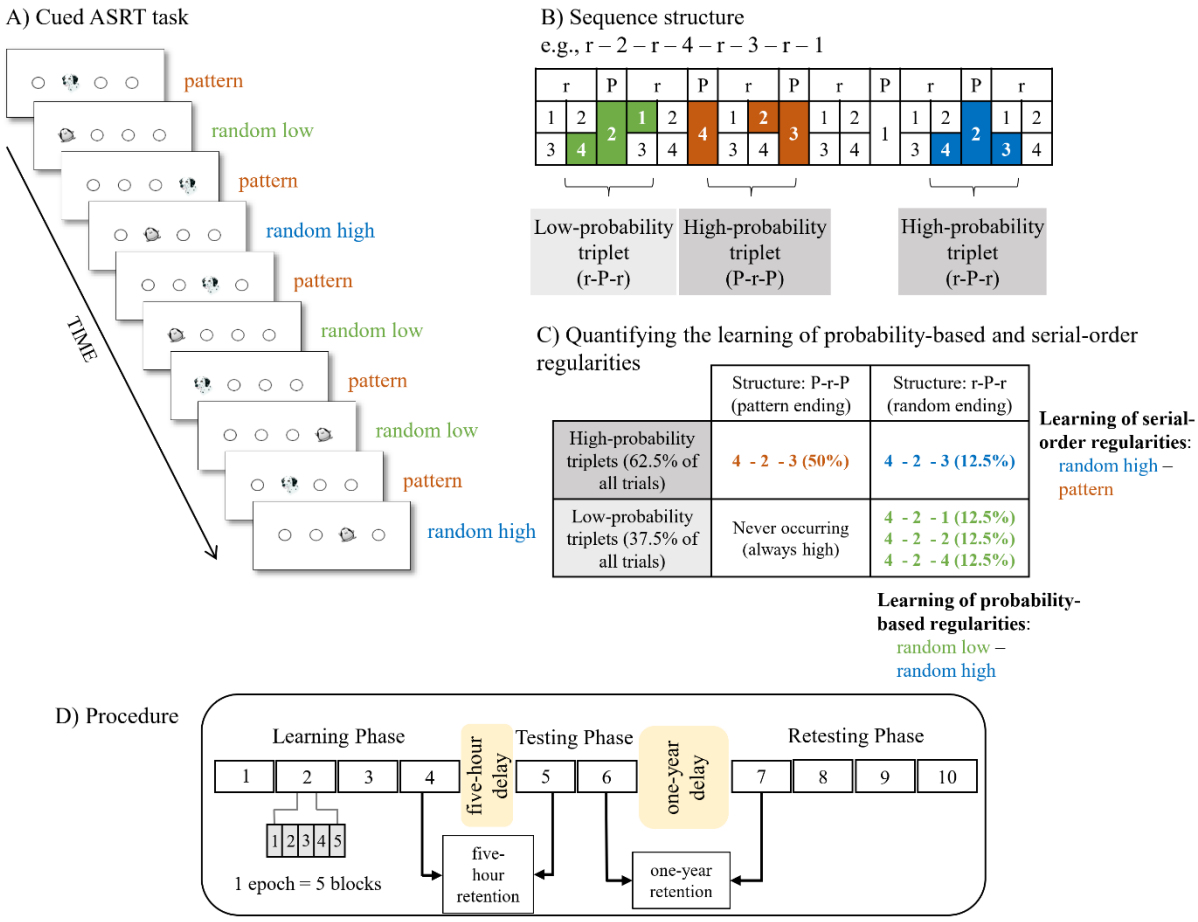

Figure 3.1. The cued Alternating Serial Reaction Time (ASRT) task and experimental procedure. (A) Pattern and random trials were presented in an alternating fashion. Pattern trials were represented by a dog's head and random trials were presented by a penguin. **(B)** An example of the sequence structure. In the example sequence, numbers mark pattern trials and 'r' marks a randomly selected location out of the four possible locations. The alternating presentation of trials makes some runs of three consecutive trials (called triplets) more probable than others, labeled as high-probability and low-probability triplets, respectively. High-probability triplets can end either with a pattern or a random trial, whereas low-probability triplets always end with a random trial. Therefore, we can differentiate pattern triplets which are always of high probability (orange shading in panel B and orange font in panel C), random high-probability triplets (blue shading in panel B and blue font in panel C), and random low-probability triplets (green shading in panel B and green font in panel C). **(C)** Quantifying the underlying learning processes in the task. Learning of probability-based regularities is quantified by contrasting the RTs on the random high and random low trials (blue vs. green, the right column of the table). Learning of serial-order regularities is calculated by contrasting the RTs on the pattern and random high trials (orange vs. blue, the top row of the table). **(D)** The design of the experiment. The experiment consisted of three sessions. The Learning Phase and Testing Phase were administered on the same day with a 5-hour offline period between them. The Learning Phase was composed of four epochs (one epoch contained 5 blocks, and each block consisted of 85 stimuli) and the Testing Phase was composed of two epochs. The four-epoch-long Retesting Phase was administered ca. one year later. Figure 3.1A, 3.1B and 3.1C are adapted from Nemeth, Janacsek, and Fiser (2013) and Zavecz et al. (2020), Figure 3.1D is adapted from Kóbor et al. (2017).

2.4. Statistical analyses

Statistical analyses were carried out by SPSS version 25.0 software and by JASP 0.9.2.0. software. We followed protocols outlined in previous studies (e.g., Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019). First, we collapsed the blocks into epochs of five blocks to facilitate data processing. The first epoch contained blocks 1-5, the second epoch contained blocks 6-10, and so forth. Hence, the Learning Phase consisted of four epochs, the Testing Phase consisted of two epochs and the Retesting Phase consisted of four epochs. Epochs are referred to consecutively (from 1 to 10).

Each trial was defined as the last element of a pattern, random high or random low triplet (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013; also see the task's description above). We calculated the median RT for correct responses separately for the three trial types, for each participant and each epoch. Based on prior studies, we excluded two types of low-probability triplets: repetitions of a single element (e.g., 111, 222) and trills (i.e., triplet starting with and ending with the same element, e.g., 121, 242) as individuals often show pre-existing tendencies towards them (Howard, Howard, Japikse, DiYanni, et al., 2004).

Based on the three trial types, learning of probability-based and serial-order regularities can be quantified (Nemeth, Janacsek, & Fiser, 2013). Learning of probability-based regularities is defined as the difference in RT between random high and random low trials (RT for random low trials minus RT for random high trials). Learning of serial-order regularities was quantified as the RT difference between pattern and random high trials (RT for random high trials minus RT for pattern trials). Hence, higher scores indicate better learning/memory of probability-based or serial-order information. In conjunction with the learning and memory scores, we also calculated an offline change score separately for knowledge of probability-based and serial-order regularities. The short-term offline change score was calculated by subtracting the memory score in Epoch 4 from the memory score in Epoch 5, therefore, it shows the change over the 5-hour offline period. The long-term offline change score was calculated by subtracting the memory score in Epoch 6 from the memory score in Epoch 7, therefore, it shows the change over the one-year offline period. In both cases, negative scores show forgetting and positive scores indicate offline learning. To assess learning and the retention of knowledge, repeated-measures ANOVAs and paired-samples *t*-tests were conducted on RT data, separately for probability-based and serial-order based regularities. The Greenhouse-Geisser epsilon (ε) correction was used when necessary. Original *df* values

and corrected *p* values (if applicable) are reported with partial eta-squared (η^2_p) as a measure of effect size. For correlation analyses, in case of normal distribution, Pearson's correlation was employed. When the assumption of normal distribution was violated, Spearman correlation was used for frequentist statistics and Kendall's Tau-b correlation was used for Bayesian statistics.

Concurrently with the frequentist analyses, Bayesian paired samples *t*-tests and independent samples *t*-tests were performed, and Bayes Factors (BF) were calculated for the relevant comparisons. The *BF* is an appropriate tool to conclude whether the data support the null (H_0) or alternative (H_1) hypothesis (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). *BF*s can be particularly relevant in memory consolidation studies where retention is indicated by evidence supporting the H_0 rather than H_1 (Dienes, 2014), as H_0 means that the memory scores before and after the offline period are similar and H_1 means that the memory scores differ. Here, we report BF_{01} values. According to Wagenmakers et al. (2011), BF_{01} values between 1 and 3 suggest anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 mean strong evidence for H_0. Values between 1 and 1/3 indicate anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 suggest strong evidence for H_1. Values around 1 do not support either hypothesis.

3. Results

3.1. Prerequisite of memory consolidation

Significant learning preceding the offline period is a prerequisite of assessing memory consolidation (Kóbor et al., 2017; Robertson, 2009). Thus, we conducted mixed-design ANOVAs on the Learning Phase to test whether significant learning of probability-based and/or serial-order regularities occurred in both the TS and TD groups. ANOVAs were conducted on RT data, separately for the two types of regularities.

Learning of probability-based regularities in the Learning Phase were tested with a mixed-design ANOVA on RT with GROUP (TS vs TD) as a between-subjects factor and PROBABILITY (random high vs random low) and EPOCH (1-4) as within-subject factors. Average RTs (i.e., irrespective of trial types) were similar in the control and TS groups (main effect of GROUP, $F(1, 36) = 0.006$, $p = .94$). RTs gradually decreased as the task progressed, irrespective of trial types (main effect of EPOCH, $F(3, 108) = 20.62$, $p < .001$, $\eta^2_p = .36$). The ANOVA revealed significant learning of

probability-based information (main effect of PROBABILITY, $F(1, 36) = 83.48$, $p < .001$, $\eta^2_p = .70$), participants showed faster responses to random high ($M = 441.20$ ms) than to random low trials ($M = 463.52$ ms). The TS and TD groups did not differ from each other either in overall learning (GROUP × PROBABILITY interaction, $F(1, 36) = 0.02$, $p = .90$; Fig. 3.2) or in the trajectory of learning (GROUP × PROBABILITY × EPOCH interaction, $F(3, 108) = 1.06$, $p = .36$). Other interactions were also not significant (all ps > .13). Successful learning in both groups ensure that the analyses of short-term and long-term consolidation of probability-based regularities across groups are justified.

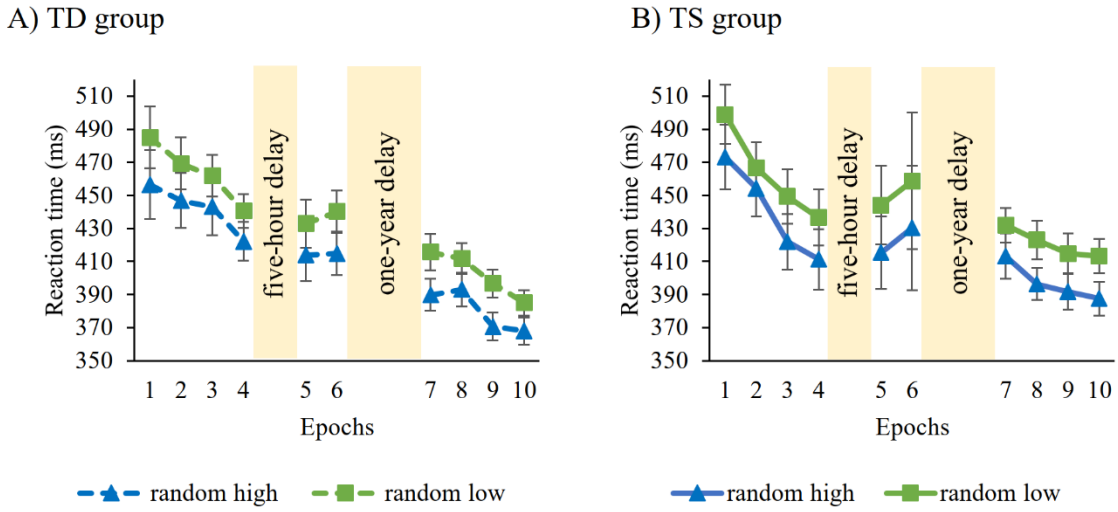

Figure 3.2. Temporal dynamics of learning of probability-based regularities across epochs and sessions in the (A) TD group and (B) TS group. Dashed lines represent the TD group, continuous lines represent the TS group. RT values as a function of the epoch (1-10) and trial types (random high vs. random low) are presented. Blue lines with triangle symbols indicate RTs on the random high trials, green lines with square symbols indicate RTs on the random low trials. Learning is quantified by the gap between blue and green lines; the greater gap between the lines represents better learning. Error bars denote standard error of mean.

Learning of serial-order regularities during the Learning Phase was tested similarly, with a mixed-design ANOVA on RT with GROUP (TS vs TD) as a between-subjects factor and ORDER (pattern vs random high) and EPOCH (1-4) as within-subject factors. Average RTs (i.e., irrespective of trials types) did not differ in the control and TS groups (main effect of GROUP, $F(1, 36) = 0.04$, $p = .85$). RTs gradually decreased as the task progressed, irrespective of trial types (main effect of EPOCH, $F(3, 108) = 28.55$, $p < .001$, $\eta^2_p = .44$). The ANOVA showed overall significant learning (main effect of ORDER, $F(1, 36) = 6.59$, $p = .015$, $\eta^2_p = .16$), participants showed faster RTs to pattern ($M = 426.47$ ms) compared to random high trials ($M = 441.20$ ms). Importantly, however, the groups differed in the trajectory of learning (indicated by the GROUP × ORDER ×

EPOCH interaction, $F(1, 36) = 5.03$, $p = .01$, $\eta^2_p = .12$, Fig. 3.3). Other interactions were not significant (all ps > .27). To further examine the three-way interaction, we investigated the learning of serial-order regularities separately in the TS and TD groups. Hence, we conducted an ANOVA on RT with ORDER (pattern vs random high) and EPOCH (1-4) separately for the two groups. In the TS group, the ANOVA did not reveal learning (non-significant main effect of ORDER, $F(1, 18) = 3.58$, $p = .08$; non-significant ORDER × EPOCH interaction, $F(3, 54) = 2.25$, $p = .09$). The TD group did not show significant learning either (non-significant main effect of ORDER, $F(1, 18) = 3.99$, $p = .06$; non-significant ORDER × EPOCH interaction, $F(3, 54) = 3.35$, $p = .07$). Importantly, these results suggest that the groups did not successfully acquire the serial-order information during the Learning Phase, therefore, the prerequisite of assessing memory consolidation was not fulfilled. The lack of significant learning calls into question the applicability of retention analyses concerning serial-order regularities. Hence, from this point on, we focus on consolidation of probability-based information and report the analysis on the consolidation of serial-order regularities in the Supplementary Materials.

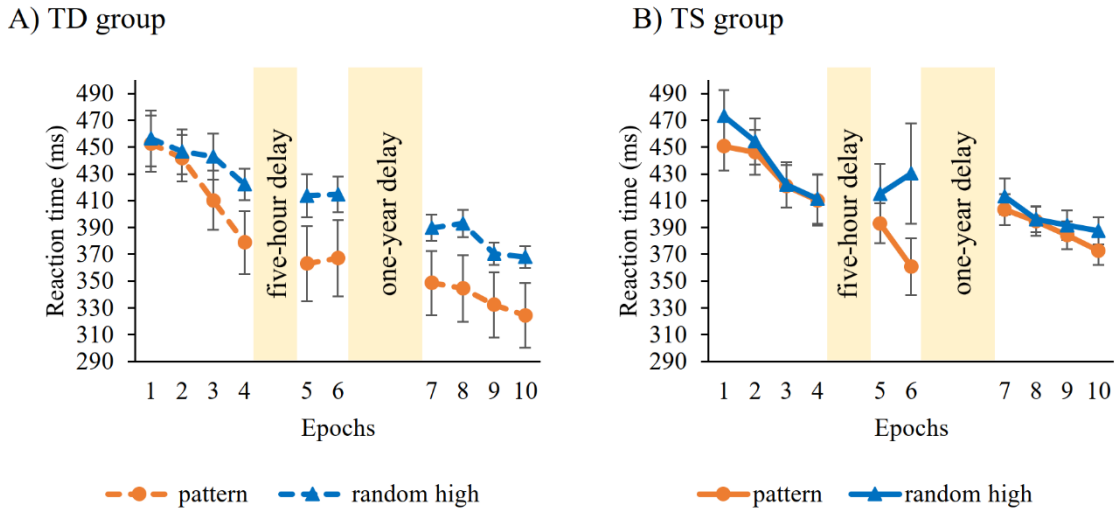

Figure 3.3. Temporal dynamics of learning of serial-order regularities across epochs and sessions in the (A) TD group and (B) TS group. Dashed lines represent the TD group, continuous lines represent the TS group. RT values as a function of the epoch (1-10) and trial types (pattern vs. random high) are presented. Orange lines with circle symbols indicate RTs on the pattern trials, blue lines with triangle symbols indicate RTs on the random high trials. Learning is quantified by the gap between orange and blue lines; the greater gap between the lines represents better learning. Error bars denote standard error of mean.

Regarding serial-order learning, we also tested the explicit knowledge of the sequence measured by the post-block sequence reports and whether it is different in the TS and control groups. Due to the violation of normal distribution, non-parametric Mann-

Whitney U test was used to contrast explicit knowledge in the TS and control groups. The two groups showed similar explicit knowledge (U = 151.5, z = -0.92, p = .36; $M_{control}$ = 79.23 %, M_{TS} = 88.42 %).

3.2. Short-term (five-hour) consolidation of knowledge of probability-based regularities

To examine the five-hour consolidation of knowledge of probability-based regularities knowledge, we conducted a mixed-design ANOVA on RT with GROUP (TS vs TD) as between-subjects factor and PROBABILITY (random high vs random low) and EPOCH (4 vs 5) as within-subject factors.

Overall, irrespective of epochs and group, participants were faster on random high (M = 415.70 ms) than on random low trials (M = 438.53 ms) (main effect of PROBABILITY, F(1, 36) = 66.37, p < .001, η^2_p = .65). The ANOVA revealed, that over groups, the memory scores did not change in the five-hour offline period (non-significant PROBABILITY × EPOCH interaction, F(1, 36) = 0.25, p = .62, BF_{01} = 5.08), with similar memory scores in the 4th (M = 21.80 ms) and in the 5th (M = 23.86 ms) epochs. Importantly, the groups did not differ in retention (non-significant GROUP × PROBABILITY × EPOCH interaction, F(1, 36) = 0.14, p = .71, Fig. 3.4; Bayesian independent samples t-tests conducted on the short-term offline change score BF_{01} = 3.004, short-term offline change scores: M_{TS} = 3.58 ms, M_{TD} = 0.53 ms). Other main effects or interactions were also not significant (all ps > .15). Furthermore, we compared the memory scores in Epoch 4 and Epoch 5 separately in the two groups with paired-samples t-tests. Both groups showed retention of probability-based regularities (TD group: t(18) = -0.08, p = .94, BF_{01} = 4.20, d = -0.02; TS group: t(18) = -0.70, p = .49, BF_{01} = 3.39, d = -0.16, see also Fig. 3.4).

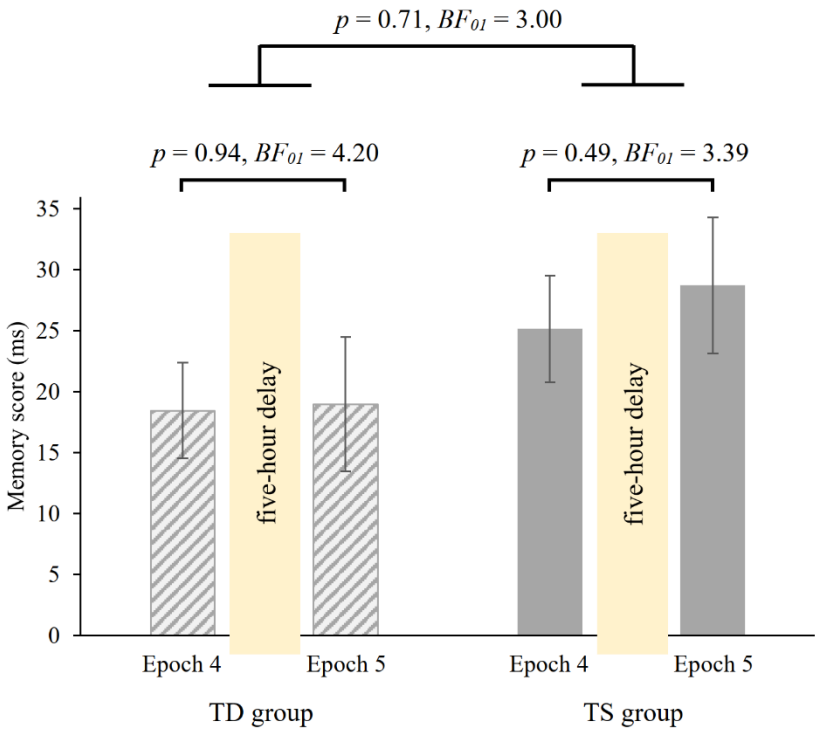

Figure 3.4. Five-hour retention of knowledge of probability-based regularities in the TD and TS groups. Memory scores were measured by RT values for the last epoch of the Learning Phase (Epoch 4) and the first epoch of the Testing Phase (Epoch 5). Error bars denote the standard error of mean.

3.3. Long-term (one-year) consolidation of knowledge of probability-based regularities

To investigate the one-year consolidation of knowledge of probability-based regularities, we run a mixed-design ANOVA on RT with GROUP (TS vs TD) as between-subjects factor and PROBABILITY (random high vs random low) and EPOCH (6 vs 7) as within-subject factors. Overall, irrespective of epochs and group, participants showed faster RTs on random high (M = 412.08 ms) than on random low trials (M = 436.68 ms) (main effect of PROBABILITY, $F(1, 36) = 87.75$, $p < .001$, $\eta^2_p = .71$). The ANOVA revealed retained memory of probability-based regularities after the one-year delay (non-significant PROBABILITY × EPOCH interaction, $F(1, 36) = 0.496$, $p = .49$, $BF_{01} = 4.53$), memory scores were similar in the 6th (M = 26.85 ms) and in the 7th (M = 22.34 ms) epochs. Importantly, memory scores were similar in the TS and TD groups (non-significant GROUP × PROBABILITY × EPOCH interaction, $F(1, 36) = 0.64$, $p = .43$, Fig. 3.5; Bayesian independent samples t-tests conducted on the long-term offline change score BF_{01} = 2.47, long-term offline change scores: M_{TS} = -9.63 ms, M_{TD} = 0.61 ms). Other main effects and interactions were also not significant (all ps > .20). Furthermore, we compared the memory scores in Epoch 6 and Epoch 7 separately in the two groups

with paired-samples *t*-tests. Both groups showed retention of probability-based regularities (TD group: $t(18) = -0.08$, $p = .93$, $BF_{01} = 4.20$, $d = -0.02$; TS group: $t(18) = 0.91$, $p = .37$, $BF_{01} = 2.92$, $d = 0.21$, see also Fig. 3.5).

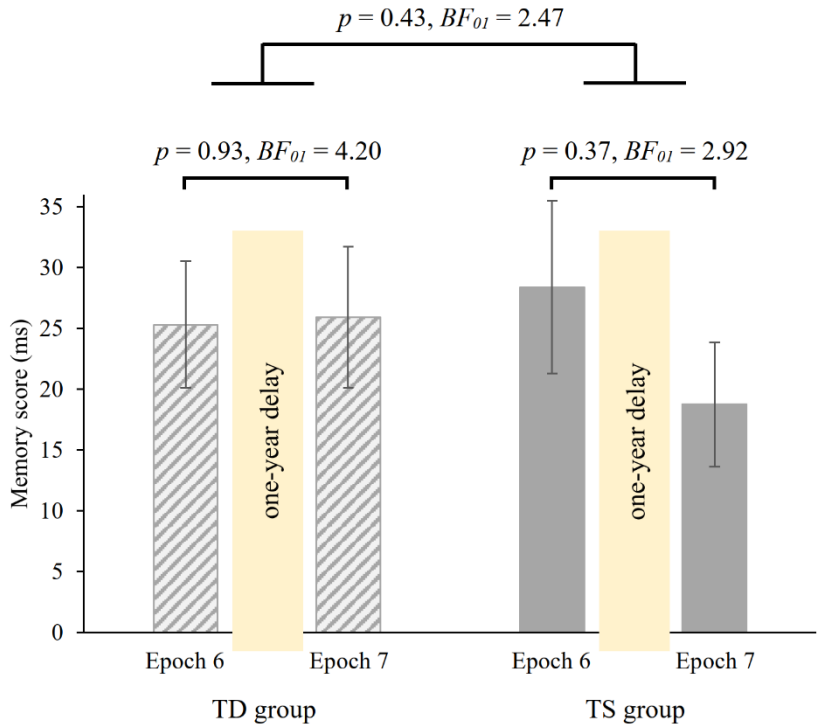

Figure 3.5. One-year retention of knowledge of probability-based regularities in the TD and TS groups. Memory scores were measured by RT values for the last epoch of the Testing Phase (Epoch 6) and the first epoch of the Retesting Phase (Epoch 7). Error bars denote the standard error of mean.

3.4. The relation of tic severity and consolidation of knowledge of probability-based regularities

In the TS group, we measured the severity of present tics on the first testing day as well as one year later, on the second testing day. This way, we could assess the change in tic severity over the one-year offline period. We subtracted the total score of tic severity on the second testing day (i.e., after the one-year offline period) from the total score of tic severity on the first testing day. Therefore, positive scores mean positive change over the one-year offline period, and negative scores mean that tics became more severe. The mean total scores on the first and second testing days are reported in Table 3.1. The mean of the change in tic severity was 0.63 ($SD = 9.55$).

To evaluate the relationship between tic severity and consolidation of knowledge of probability-based regularities, we correlated short- and long-term offline change scores

and tic severity on the first testing day. Severity of the present tics on the first testing day did not correlate either with short-term offline change score ($rs(19) = -.10$, $p = .68$, $BF_{01} = 3.22$), or with long-term offline change score ($rs(19) = -.07$, $p = .77$, $BF_{01} = 3.32$). Moreover, we correlated the long-term offline change score of knowledge of probability-based regularities with the change in tic severity over the one-year offline period and found no correlation between the variables ($rs(19) = -.19$, $p = .43$, $BF_{01} = 2.57$).

4. Discussion

The present study aimed to investigate the short-term (five-hour) and long-term (one-year) consolidation of two aspects of procedural memory, namely probability-based and serial-order regularities, in children with Tourette syndrome and neurotypical peers. We employed the cued ASRT task, which measures the two aspects simultaneously. We have shown retained knowledge of probability-based information: participants acquired the probability-based regularities, then successfully retained them both after the five-hour and one-year offline period. Children with TS and matched typically developing controls showed comparable retention of knowledge of probability-based regularities. These results were supported by Bayesian statistics as well, strengthening the evidence for successful five-hour and one-year retention in both groups. Concerning serial-order regularities, the prerequisite of assessing memory consolidation was not fulfilled as the groups did not acquire the serial-order information. Hence, consolidation of serial-order information could not be reliably tested here. Nevertheless, we presented these results in the Supplementary Material showing successful retention in both groups.

Previous studies already demonstrated retained memory representation of probability-based information in neurotypical adults following a one-year offline period using the ASRT task (Kóbor et al., 2017; Romano et al., 2010). Importantly, evidence for successful retention was presented in neurotypical children as well in the study of Tóth-Fáber, Janacsek, et al. (2021). Altogether, these studies suggest that one-year consolidation of probability-based regularities seems to be comparable between children and adults, supporting the age invariance model in the consolidation of such regularities. Our results corroborate these prior findings and also demonstrates intact one-year consolidation of probability-based information in children with TS, suggesting that procedural memory is robust in this neurodevelopmental disorder.

The intact consolidation of knowledge of probability-based regularities in TS is in line with the results of prior studies. Takács et al. (2018) employed the uncued version of the ASRT task and probed learning and consolidation of probability-based regularities in TS. Children with TS showed superior learning, however, following a 16-hour offline period, they showed greater forgetting than their neurotypical peers. Importantly, group differences in learning itself can lead to group differences in consolidation. When controlling for learning differences, the TS and control group showed similar changes in knowledge of probability-based regularities overnight, suggesting that consolidation is comparable between the groups. The present study replicates and goes beyond the results of Takács et al. (2018): (1) our results also showed comparable short-term (five-hour) consolidation, and (2) we showed intact one-year retention of knowledge of probability-based regularities in TS.

Consolidation of procedural memory representations in TS has also been indirectly examined with language-based tasks that measure the access to previously established procedural information. Walenski et al. (2007) showed faster production of rule-governed past tenses and faster naming of manipulated objects in TS, whereas production of irregular past tenses and naming non-manipulated objects were similar between the TS and typically developing groups. The former processes rely on procedural memory and the latter processes are related to declarative memory (Ullman, 2004). In conjunction with these results, Dye et al. (2016) found evidence for enhanced access to established procedural information in TS in a non-word repetition task: children with TS could repeat non-words in a faster manner than typically developing controls. In sum, the results of both Walenski et al. (2007) and Dye et al. (2016) suggested enhanced access to previously consolidated procedural information in TS, whereas our study showed intact procedural memory consolidation. This discrepancy might be explained by the differences between the employed tasks. Although processing of probability-based regularities is important in language (e.g., Misyak et al., 2010; Saffran et al., 1996; Thompson & Newport, 2007), the ASRT task employed in the current study and the language-based tasks used in the prior studies (Dye et al., 2016; Walenski et al., 2007) show some differences. Both language-based studies used stimuli or measured processes participants have a prior knowledge of: Walenski et al. (2007) used words as stimuli and Dye et al. (2016) measured rule-governed (de)composition of words. In contrast, participants had no prior knowledge of the stimuli and the underlying structure presented in the ASRT task. Moreover, participants have a repeated exposure to the stimuli in the

language-based tasks besides the experimental sessions, hence, the stimuli are retained, and therefore the memory representations are reinforced regularly, whereas participants met with the ASRT task solely in the experimental sessions. Speculatively, it is possible that regular practice in the offline period is needed for the enhancement seen in the language-based studies (Dye et al., 2016; Walenski et al., 2007). Further studies are warranted to test this possibility by investigating language-related processes and extraction of regularities during the lifespan of TS patients.

Long-term stability of procedural memories has potential clinical and educational implications. Procedural memory underlies the acquisition of cognitive, motor and social skills, such as language learning or sports (Frith & Frith, 2012; Kaufman et al., 2010) and is related to habits as well (Goodman et al., 2014; Takacs et al., 2021). Importantly, a one-year long offline period that has been used in the current study can resemble real-word observations. Namely, learning a new skill or developing a habit happens over a longer stretch of time than a timescale of a lab visit. Our results suggest that children with TS have stable memory representations of procedural knowledge without additional practice during a long time interval and their performance is comparable with TD children. These robust memory representations of procedural knowledge could manifest in everyday settings in the following way: children with TS might be better in learning a new skill (as suggested by prior studies on procedural learning, Takács et al., 2018; Tóth-Fáber, Tárnok, et al., 2021) and they can be also successful in maintaining and remembering those skills, as the current study suggests. As for clinical settings, behavioral therapies are first-line treatments for reducing tics. This method can potentially benefit from a dovetailed knowledge of how stable the acquired skills are in TS. For instance, in habit reversal training (Piacentini & Chang, 2005), when feeling the urge to tic, patients learn to perform an adequate, antagonist action that is physically incompatible with the tic. Over time and practice, when feeling the urge, patients will carry out the adequate action instead of the tic, hence, the urge – tic association will be replaced by the urge – adequate action association. It is conceivable that the new association results in a stable memory representation just as tics and procedural knowledge. Relatedly, Petruo et al. (2020) examined the effect of behavioral therapy on procedural associations and inhibitory control in adolescents with TS and typically developing peers. They employed the Comprehensive Behavioral Intervention for Tics (CBIT, Piacentini et al., 2010), which is a complex behavioral therapy consisting of psychoeducation, relaxation and habit reversal training. Procedural associations and

inhibitory control were tested preceding and following the intervention. Participants with TS showed worse inhibitory control than TD peers during the pretest, however, performance was comparable between the groups during the posttest. These results suggest that CBIT reduced the rigid procedural associations in TS, resulting in successful inhibitory control. The authors concluded that, in conjunction with reducing tics, CBIT/HRT might normalize alterations in higher level cognitive function (Petruo et al., 2020). Furthermore, one intriguing but so far neglected area which has great clinical implications is the rewiring of memory representations (Szegedi-Hallgató et al., 2017) in TS. Previous studies have suggested that memory representations of procedural knowledge might be overstable in TS (Shephard et al., 2019; Takacs et al., 2021), which would result in higher resistance to interference. This would lead to more difficulties in rewiring/overwriting established associations. Future studies should test how well patients with TS can overwrite associations and how it relates to behavioral therapies. In sum, it is important to note that the abovementioned clinical implications are tentative and future studies are necessary to investigate the relationship between procedural functions and behavioral therapies in TS.

The present study is not without limitations. First, the sample size in our study can be considered to be small. At the same time, this sample size corresponds to previous studies that investigated procedural functions in this rare disorder (e.g., Dye et al., 2016; Shephard et al., 2019; Takács et al., 2018; Takács et al., 2017). Second, TS participants in our study are characterized with mild to moderate tic severity and possibly represent the lower end of tic severity dimension as indicated by the YGTSS. In our study, individual differences in procedural memory consolidation did not correlate with tic severity or changes in tic severity over the one-year offline period in the TS group. Although YGTSS is a well-established clinical measure of tic severity, it would be beneficial to employ additional measures of tic severity in future studies. One potential candidate is the Modified Rush Videotape Rating Scale (Goetz, Pappert, Louis, Raman, & Leurgans, 1999), which is an objective clinical measure of present tic severity. Vocal and motor tics are rated based on their frequency, severity, and distribution. Kleimaker et al. (2020) focused on stimulus-response associations in TS adults and employed both the YGTSS and the Modified Rush Videotape Rating Scale. They revealed stronger stimulus-response binding in TS adults, which resulted in difficulties in unbinding and rebinding established associations. Individual differences in stimulus-response binding did not show correlation with YGTSS scores, however, they correlated with motor tic frequency,

that is, with the number of motor tics per minute as measured by the Modified Rush Videotape Rating Scale. Hence, it is possible that finer markers of tics might be related to procedural memory consolidation as well. Nonetheless, our results are limited to a specific TS population and further studies should investigate whether procedural memory consolidation is intact in children with severe tic symptoms as well.

Another intriguing future direction could be a detailed characterization of the temporal dynamics of consolidation. The present study does not provide information about exactly when the consolidation of the acquired information happens. Quentin et al. (2021) showed that learning of probability-based regularities occurs during the online periods, and no further gain is acquired during the offline periods. In contrast, memories of serial-order regularities are formed during the offline periods of the ASRT task. Further studies should explore in detail the temporal dynamics of learning probability-based and serial-order regularities in TS.

5. Conclusion

The goal of the present study was to investigate the consolidation of procedural memory in TS. The representation of probability-based regularities remained stable over both a short-term (five-hour) and long-term (one-year) offline period in children with TS and typically developing controls. Both the TS and the control group successfully retained knowledge of probability-based information after the offline periods with comparable memory performance between the groups. In conclusion, procedural memory consolidation seems to be intact in TS even after a one-year offline period that did not include additional practice. This finding suggests that individuals with TS might be more proficient in skill acquisition as they are able to successfully maintain and retain the learned skills, even over a long period of time.

IV. Study 3: Statistical and sequence learning lead to persistent memory in children after a one-year offline period

Publication:

Tóth-Fáber, E., Janacsek, K. & Nemeth, D. (2021). Statistical and sequence learning lead to persistent memory in children after a one-year offline period. *Scientific Reports, 11,* 12418. https://doi.org/10.1038/s41598-021-90560-5

Abstract

Extraction of environmental patterns underlies human learning throughout the lifespan and plays a crucial role not only in cognitive but also perceptual, motor, and social skills. At least two types of regularities contribute to acquiring skills: (1) statistical, probability-based regularities, and (2) serial order-based regularities. Memory performance of probability-based and/or serial order-based regularities over short periods (from minutes to weeks) has been widely investigated across the lifespan. However, long-term (months or year-long) memory performance of such knowledge has received relatively less attention and has not been assessed in children yet. Here, we aimed to test the long-term memory performance of probability-based and serial order-based regularities over a one-year offline period in neurotypical children between the age of 9 and 15. Participants performed a visuomotor four-choice reaction time task designed to measure the acquisition of probability-based and serial order-based regularities simultaneously. Short-term consolidation effects were controlled by retesting their performance after a 5-hour delay. They were then retested on the same task one year later without any practice between the sessions. Participants successfully acquired both probability-based and serial order-based regularities and retained both types of knowledge over the one-year period. The successful retention was independent of age. Our study demonstrates that the representation of probability-based and serial order-based regularities remains stable over a long period of time. These findings offer indirect evidence for the developmental invariance model of skill consolidation.

1. Introduction

Detecting and extracting various kinds of regularities embedded in our environment is a fundamental component underlying human learning in all ages, enabling us to adapt to our surroundings and to predict future events (Conway, 2020; Frost, Armstrong, Siegelman, & Christiansen, 2015; Misyak et al., 2010; Saffran & Kirkham, 2018). Extraction of regularities is argued to be the basis of several motor and cognitive skills, including language (Arciuli & Torkildsen, 2012; Conway, 2020; Kaufman et al., 2010; Saffran et al., 1996; Siegelman, 2020; Ullman, 2004). The initially unstable representations of the detected and extracted regularities are converted into a more stable form via consolidation, allowing information to be preserved and retained later (Walker, 2005). Ample studies investigated the consolidation of regularities and skills with a one-minute, one-hour, four-hour, 12-hour, 24-hour or one-week delay (Arciuli & Simpson, 2012; Fanuel et al., 2020; Meier & Cock, 2014; Nemeth & Janacsek, 2011; Nemeth, Janacsek, Londe, et al., 2010; Press, Casement, Pascual-Leone, & Robertson, 2005; Simor et al., 2019; Song et al., 2007b; Walker, Brakefield, Hobson, & Stickgold, 2003). Although everyday experiences suggest that the representation of the acquired regularities and skills is persistent even for a more extended period (months or years), it has been rarely tested empirically, especially from a developmental perspective. In the present study, we aim to investigate the long-term (one-year) consolidation of two types of regularities in neurotypical children.

Behaviorally, consolidation is measured by contrasting memory performance at the end of the learning session with performance at the beginning of a subsequent testing session, without additional practice between the two sessions (i.e., during the offline periods). Consolidation can be expressed by successfully retained knowledge after the offline period (no forgetting, i.e., performance is similar in the learning and testing sessions) or by learning-dependent, delayed performance gains after the offline period, termed offline learning (i.e., performance is better in the testing session than in the learning session) (Robertson et al., 2004). The present study follows this well-established behavioral test protocol to assess long-term memory performance and implements a one-year offline period between the sessions.

The available information in our environment, which can be detected, extracted, and consolidated is diverse; thus, our brain has to process several information streams simultaneously during both learning and consolidation. Learning of regularities is not a

monolithic process. Previous empirical studies suggested that the acquisition of at least two types of regularities can be differentiated: (1) statistical, probability-based regularities, and (2) serial order-based regularities (Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019; Takacs et al., 2020). Therefore, it has been proposed that humans organize the regularities embedded in the environment in separate hypothesis spaces (Conway, 2020; Maheu et al., 2020); one such hypothesis space is based on probabilities, while another is based on deterministic rules (i.e., serial order-based regularities). While on the level of transitional probabilities, these learning processes may seem highly similar, where the former can be viewed as the acquisition of transitional probabilities that are less than one and the latter as the acquisition of transitional probabilities that equal one, differences between them have been shown both on the behavioral and neural levels, providing support for their distinction(Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019). Probability-based regularities are picked up rapidly, while learning of serial order-based information follows a more gradual trajectory (Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019); they also manifest differently on the level of event-related potentials (Kóbor et al., 2018; Takacs et al., 2020) and show different neural oscillations during consolidation (Simor et al., 2019; Zavecz et al., 2020).

The consolidation of probability-based and/or serial order-based regularities has been studied previously. Retained knowledge has been found after one-hour, 12-hour, 24-hour, or even 1-week offline period in healthy adults (Arciuli & Simpson, 2012; Meier & Cock, 2014; Nemeth & Janacsek, 2011; Nemeth, Janacsek, Londe, et al., 2010; Simor et al., 2019; Song et al., 2007b). Long-term consolidation has received less attention, with only a few studies investigating the effect of month- or year-long offline periods: Romano et al. (2010) and Kóbor et al. (2017) both showed persistent representation of regularities after a one-year offline period in healthy adults. However, both studies employed a task (Kóbor et al., 2017; Romano et al., 2010), which, although measures both probability-based and serial order-based information, is not well-suited to dissect these regularities in the same time window. To the best of our knowledge, only two studies investigated the consolidation differences between probability-based and serial order-based regularities, but they administered a one-hour offline period only (Simor et al., 2019; Zavecz et al., 2020). Both Simor et al. (2019) and Zavecz et al. (2020) found retained statistical and serial-order knowledge after the offline period. Altogether, prior studies typically incorporated only short-term (from minutes to week) offline periods in their design;

therefore, the long-term consolidation of probability-based and serial order-based information is not well understood yet. Here, we aimed to fill this gap by investigating the simultaneous consolidation of these regularities over a one-year period.

The consolidation of probability-based or serial order-based information is even less understood in children. An ideal avenue to achieve a deeper understanding of cognitive processes and functions is to examine them from a developmental perspective (Karmiloff-Smith, 1994). Studies on typical and atypical development can pave the way towards grasping underlying processes of learning and memory consolidation. Learning of probability-based regularities might be age-variant with better performance in children up to the age of 12 (Janacsek et al., 2012; Juhasz et al., 2019; Nemeth, Janacsek, & Fiser, 2013), whereas learning of serial order-based regularities might be comparable in children and adults (Nemeth, Janacsek, & Fiser, 2013). Most of the studies examining the consolidation of these regularities in children either focused on solely probability-based (Fischer et al., 2007) or solely serial order-based regularities (Desmottes, Maillart, & Meulemans, 2017; Hedenius et al., 2013) or used paradigms that intermix them (Hedenius, Lum, & Bölte, 2020; Hedenius et al., 2011; Nemeth, Janacsek, Balogh, et al., 2010; Takács et al., 2018). Retained information (i.e., no forgetting) has been found in neurotypical children following 11-hour (Fischer et al., 2007), 16-hour (Nemeth, Janacsek, Balogh, et al., 2010; Takács et al., 2018), 24-hour (Hedenius et al., 2013) and 3-day (Hedenius et al., 2011) offline periods. Hedenius et al. (2020) showed offline learning after a 24-hour delay and Desmottes et al. (2017) found offline learning following 24-hour and one-week offline periods. To the best of our knowledge, the long-term (one-year) consolidation of probability-based or serial order-based information has not yet been investigated in children. Age-variant learning of probability-based regularities and successful one-year retention in healthy adults (Kóbor et al., 2017; Romano et al., 2010) raises the question of whether long-term consolidation is successful in children as well.

To sum up, in child population, the long-term (one-year) consolidation of probability-based and serial order-based information has not been assessed yet. In the present study, we used the cued version of the Alternating Serial Reaction Time task (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013) which enables us to simultaneously measure these two regularities. In this framework, statistical learning refers to the acquisition of short-range, temporally distributed probability-based information between visual stimuli. Sequence learning refers to the acquisition of serial

order-based information, where participants are exposed to stimuli that repeatedly occur in the same deterministic order, incorporated with random stimuli (hence, creating an alternating sequence structure). Our study aims to examine one-year consolidation of probability-based and serial order-based regularities in children between the age of 9 and 15 with a task designed to measure the acquisition of these regularities simultaneously. This particular age range was chosen in order to examine consolidation both in childhood and adolescence, hence, participants from pre-adolescence, early- and mid-adolescence were included. Based on the previous studies, we expect successful retention of both probability-based and serial order-based information following a one-year offline period.

2. Methods

2.1. Participants

Seventy-eight children between the age of 9 and 15 participated in our study from local schools. Three participant had missing data on the ASRT task due to technical difficulties; three children's caregiver reported psychiatric condition; one child did not have corrected-to-normal vision during one session of the assessments; and one children showed extremely low average accuracy according to Tukey (1977) criterion (more than 3 times the interquartile range) consistently throughout the ASRT task. These eight participants were excluded from the analyses. The final sample consisted of 70 participants (M_{age} = 11.99 years, SD_{age} = 1.61 years; 37 boys, 33 girls).

Participants performed in the normal range on standard neuropsychological tests (Wisconsin Card Sorting Task (Berg, 1948; Piper et al., 2015) [WCST, percentage of perseverative errors]: *M* = 14.12%, *SD* = 5.90%; Counting Span task (Case et al., 1982; Fekete et al., 2010): *M* = 3.17, *SD* = 0.78). Due to technical problems, data of three participants on the WCST and data of one participant on the WCST and Counting Span task is missing. Handedness was measured by the Edinburgh Handedness Inventory (Oldfield, 1971) (EHI). Due to a technical error, EHI of one participant is missing. The Laterality Quotient (LQ) of the sample varied between -100 and 100 (where -100 means compete left-handedness and 100 means complete right-handedness) with a mean of 76.21 (*SD* = 36.50).

Furthermore, caregivers of participants completed a parental questionnaire regarding socioeconomic status (SES) and health-related questions (i.e., whether the child has any neurological, psychiatric, or neurodevelopmental disorder). Caregivers of one

participant did not provide information about their socioeconomic status (SES), therefore, data of one participant is missing. SES was determined by how many years the caregivers spent in formal education. We calculated the caregivers' average formal education based on both parents' education. In case of five participants, we only had information about one caregiver. The range of formal education of caregivers was between 9.5 and 27 years, with a mean of 16.61 years, (*SD* = 3.85 years). Caregivers were also asked to fill in the Strength and Difficulties Questionnaire (Goodman, 1997) (SDQ) which measures hyperactivity, conduct problems, emotional problems, and difficulties in peer relationships. SDQ of six participants is missing. Total problem score measured in our sample was 7.94 (*SD* = 5.38), which is well in the normal range of typically developing children (Turi, Tóth, & Gervai, 2011). All participants in the final sample had normal or corrected-to-normal vision, and none of the children had any neurological, psychiatric, or neurodevelopmental disorders according to parental reports.

Caregivers of all participants provided informed written consent, and children provided informed verbal consent to participate in the study before enrollment. The study was approved by the research ethics committee of Eötvös Loránd University, Budapest, Hungary (2018/239), and was conducted in accordance with the Declaration of Helsinki.

2.2. Task

The detection and extraction of probability-based and serial order-based regularities was measured by the cued version of the Alternating Serial Reaction Time (ASRT) task (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013). In this task, four equally spaced, horizontally arranged empty circles were presented on the screen, and a stimulus (either a dog's head or a penguin) appeared in one of the possible locations (i.e., in the empty circles) (Fig. 4.1A). The task was bimanual and participants were asked to press the corresponding key as accurately and as fast as they could using the index and middle fingers of both hands. After the response of the participant, the next target appeared 120 ms later.

The presentation of the stimuli followed an eight-element alternating sequence within which pattern and random elements alternated with each other (e.g., 1-r-2-r-4-r-3-r, where numbers indicate the locations from left to right and 'r' indicates a randomly selected location out of the four possible ones). In the cued ASRT task, pattern and random elements are marked by different visual stimuli, where pattern elements are denoted by the dog's head, and random elements are indicated by the penguins.

Participants were informed about the *presence* of the sequence and about the fact that the appearance of dogs always follows a predetermined pattern, while penguins always appear in random order. They were *not* informed about the exact sequence; they were instructed to find the pattern of the dogs' appearance to improve their performance. The alternating sequence makes six different sequence variations: 1-r-2-r-3-r-4-r, 1-r-2-r-4-r-3-r, 1-r-3-r-2-r-4-r, 1-r-3-r-4-r-2-r, 1-r-4-r-2-r-3-r, and 1-r-4-r-3-r-2-r. These permutations can start at any location (e.g., 1-r-2-r-3-r-4-r and 2-r-3-r-4-r-1-r are identical sequence permutations). One of the permutations was selected for each participant in a counterbalanced fashion across participants. For a given participant, the permutation remained the same through all sessions. The stimuli were presented in blocks, each block consisting of 85 trials. Each block started with five random trials for practice, then one of the eight-element alternating sequence was presented ten times.

In the task, three successive trials are referred to as triplets. Due to the alternating sequence in the task, some triplets are more probable than others. Each trial is categorized as the last element of a triplet in a moving window manner, which means that a given trial is characterized as the first element of a given triplet, as the second element of the following triplet and also as the last element of the next triplet, irrespective of whether it is a pattern or random trial. In the example sequence of Figure 4.1B, 2-r-4-r-3-r-1-r (numbers indicate the locations from left to right and 'r' indicates a randomly selected location out of the four possible ones), 2-X-4, 4-X-3, 3-X-1 and 1-X-2 (where X represents the middle element of the triplet) appeared with a higher probability because their third element could have been either pattern or random. Note that here, we use X to indicate the middle element of the triplet because for example 4-X-3 can appear both as 4-2-3 (Pattern – random – Pattern) where the first and last elements are part of the predetermined pattern and as 4-2-3 (random – Pattern – random) where the first and last elements are random, and the middle element is part of the predetermined pattern (see also Fig. 4.1B and 4.1C). In contrast, triplets such as 4-2-1 or 4-2-2 occurred with a lower probability because their third element could have been only random (that is, random – Pattern – random structure). The former triplet types are called "high-probability" triplets, while the latter types are labeled as "low-probability" triplets (Howard, & Howard, 1997).

Besides probability, another important aspect of the elements is their structure, meaning whether they are pattern or random elements. High-probability triplets can be differentiated based on their last element being a pattern element or a random element. The third element of low-probability triplets can only be random since pattern elements

always appear with high probability. Importantly, performance is not operationalized on the level of triplets; instead, performance (i.e., reaction time) is always calculated only on the last element of a triplet. Each element (i.e., trial) was categorized as the third element of either a high-probability or a low-probability triplet and also either as pattern or random elements (note that they are also visually distinguishable).

There are 64 unique triplets in the task, including all Pattern – random – Pattern (50%) and random – Pattern – random (50%) triplets; 16 triplets are high-probability triplets, and 48 triplets are low-probability ones. Regarding high-probability triplets, there are four possible combinations in regard to the first and third elements of the triplet (for the example sequence: 2-r-4-r-3-r-1-r, 2-X-4, 4-X-3, 3-X-1 and 1-X-2) with four possible locations for the middle element. In detail, the high-probability triplet of 4-X-3 can be 4-1-3, 4-2-3, 4-3-3 and 4-4-3. Since high-probability triplets can occur as Pattern – random – Pattern (50%) and by 1/4 chance as random – Pattern – random (12.5%), these triplets constitute 62.5% of all trials (Fig. 4.1C). As for low-probability triplets, the first and the second element of the triplet can appear on any of the four locations, whereas the last element has three possible locations as the fourth one corresponds to a high-probability triplet. Thus, low-probability triplets constitute 37.5% of all trials. As noted above, all low-probability triplets have a random – Pattern – random structure. On the level of unique triplets, high-probability triplets are five times more probable than the low-probability ones (4% [62.5%/16] vs. 0.8% [37.5%/48]).

Altogether, three trial types can be distinguished: (1) trials that belong to the predetermined sequence and are the last element of a high-probability triplet labeled as *pattern* trials (such as 4-2-3 in Fig. 4.1B and 4.1C marked with orange); (2) random elements that are the last element of a high-probability triplet called random high trials (such as 4-2-3 in Fig. 4.1B and 4.1C marked with blue); and (3) random elements that are the last element of a low-probability triplet labeled as *random low trials* (such as 4-2-1 in Fig. 4.1B and 4.1C marked with green).

Prior studies have demonstrated that participants show gradually faster responses to high-probability triplets than low-probability ones as the task progresses (i.e., triplet learning in the original, uncued version of the ASRT task). However, as high-probability triplets consist of both pattern and random triplets, serial-order knowledge cannot be measured by comparing merely the high- and low-probability triplets. It is important to note that in the uncued ASRT task, the underlying structure is identical to the one in the cued ASRT task, however, pattern and random stimuli are not visually distinguishable

(Howard, & Howard, 1997). Due to the identical underlying structure, the dissection of probability-based and serial-order based regularities is possible in the uncued ASRT task, but excessive training (i.e., 4-day-long training) is needed for the acquisition of serial-order knowledge (Howard, Howard, Japikse, DiYanni, et al., 2004; Howard, & Howard, 1997). The visual distinction of pattern and random trials ensures that the acquisition of both probability-based and serial-order based regularities – also referred to as statistical and sequence learning, respectively (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013) – can be measured within the same time frame.

Statistical learning is quantified by the difference in accuracy or reaction times (RTs) between random high- and random low-probability trials. These trials share the same sequence properties as they are both random but differ in statistical properties as, on the level of unique triplets, the former ones are more probable than the latter ones (see details above). Sequence learning is measured by the difference in accuracy or RTs between pattern and random high-probability trials. The statistical properties of these trials are identical as they are both highly probable, while their sequence properties differ as pattern trials are part of the predetermined sequence. In conclusion, statistical learning refers to the acquisition of purely probability-based information, while sequence learning captures the acquisition of serial order-based information (Fig. 4.1C).

2.3. Procedure

The experiment was composed of three sessions. The first two sessions were administered on the same day with a 5-hour delay between them, while the third session was administered ca. one year later (M_{delay} = 53.08 weeks, SD_{delay} = 2.39 weeks, between 47.95 and 60.24 weeks, Fig. 4.1D). The ASRT task was administered in all three sessions. In the Learning Phase, participants completed 20 blocks, which, during the statistical analyses, were collapsed into epochs, each containing five blocks. The Testing Phase consisted of 10 blocks (i.e., two epochs), while the Retesting Phase contained 20 blocks (i.e., four epochs) (Fig. 4.1D). Participants were assessed in a quiet room in their school. During the 5-hour offline period on the first day, they continued with their school activities such as classes and extracurricular activities. At the end of the first day (i.e., after the Learning and Testing Phases), participants were not informed that they would perform the task one year later.

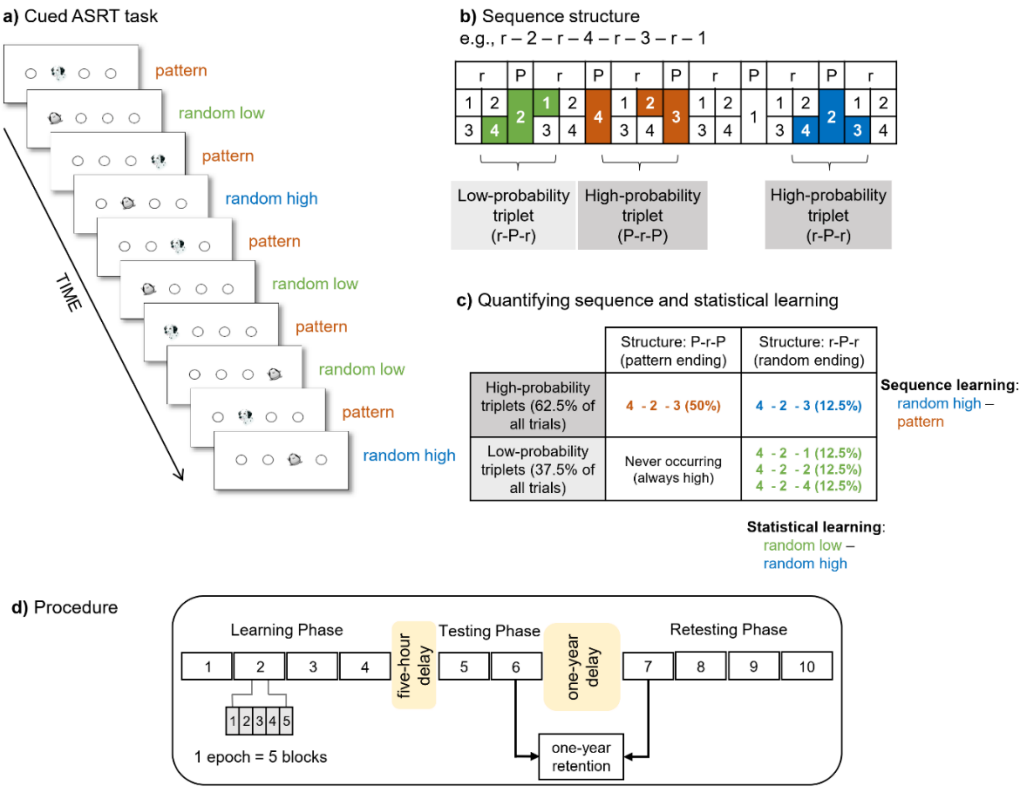

Figure 4.1. The cued Alternating Serial Reaction (ASRT) task and experimental procedure. **(A)** Pattern and random trials were presented in an alternating fashion. Pattern trials were marked by a picture of a dog's head, and random trials were marked by a picture of a penguin. **(B)** An example of the sequence structure. Numbers indicate pattern trials, and 'r' indicates a randomly selected location out of the four possible ones. The alternating sequence makes some runs of three consecutive trials (labeled as triplets) more probable than others, called high-probability and low-probability triplets, respectively. Among high-probability triplets, the last element of the triplet can be either pattern or random. Based on this, we could determine pattern high-probability triplets (with orange shading in panel B and orange font in panel C) and random high-probability triplets (with blue shading in panel B and blue font in panel C). Among low-probability triplets, only random low-probability triplets can occur (with green shading in panel B and green font in panel C). **(C)** The underlying learning processes measured by the task. Statistical learning is calculated by contrasting the accuracies or RTs on the random high and random low trials (blue vs. green, the right column of the table). Sequence learning is quantified by contrasting the accuracies or RTs on the pattern and random high trials (orange vs. blue, the top row of the table). The table presents the calculation of learning processes on RT data. **(D)** The design of the experiment. The experiment was composed of three sessions. The Learning Phase consisted of four epochs (one epoch contained 5 blocks, and each block consisted of 85 stimuli), followed by a 5-hour offline period then the two-epoch-long Testing Phase on the same day. The Retesting Phase with four epochs was administered ca. one year later. Figure 4.1A, 4.1B, and 4.1C are adapted from Nemeth, Janacsek, and Fiser (2013), Figure 4.1D is adapted from Kóbor et al. (2017).

2.4. Statistical analyses

Statistical analyses were carried out by SPSS version 25.0 software (SPSS, IBM) and by JASP 0.11.1.0. software (JASP, 2019). Based on previous studies using the ASRT task (Kóbor et al., 2018; Nemeth, Janacsek, & Fiser, 2013; Simor et al., 2019), we firstly collapsed the blocks of the task into epochs, with each epoch consisting of five blocks.

This way, the Learning Phase contained four epochs, the Testing Phase contained two epochs, while the Retesting Phase consisted of four epochs. Epochs are labeled consecutively (from 1 to 10, Fig. 4.1D). From the analysis, two types of low-probability triplets were excluded: repetitions (e.g., 111, 222) and trills (e.g., 121, 242), as participants often show pre-existing tendencies to them (Song et al., 2007a, 2007b). As described above in the task description, each trial was determined as the last trial of a pattern, random high, or random low triplet. Mean accuracy (ratio of correct responses) and median RT (for correct responses) were calculated for each participant and each epoch, separately for the three types of trials (i.e., pattern, random high and random low trials). Based on the three trial types, statistical and sequence learning can be assessed by the cued ASRT task (Nemeth, Janacsek, & Fiser, 2013) (for further details, see the task's description above). Analyses and results concerning accuracy are presented in the Supplementary Material; here, we focus on RT data.

Prior developmental studies showed that age has a large effect on average RTs, with younger children showing slower RTs (e.g., Janacsek et al., 2012; Juhasz et al., 2019; Zwart et al., 2019). To test this, we first calculated average RTs over the 10 epochs (i.e., RT data were calculated on all correct trials, irrespective of trial types). We then correlated the average RTs with age: the analysis revealed significant negative correlation ($r(68) = -.54$, $p < .001$), showing that younger children were slower on the task. To control for the effect of average RT differences related to age on learning and consolidation of knowledge, we transformed the data in the following way. We divided each participants' raw RT values of each trial type and each epoch by their own average performance (i.e., average RT) in the first epoch of the task (for a similar approach, see Horvath et al., 2020; Nitsche et al., 2003). Participants' performance was around 1 at the beginning of the task and changed as the task progressed. Values above 1 meant that responses were slower on a given trial type than average RTs in the very first epoch of the task, while values below 1 indicated faster responses on a given trial type compared to average RTs in the first epoch. We conducted all analyses on standardized RT data.

Statistical learning score in the Learning Phase and memory scores in the Testing and Retesting Phases were quantified as the difference between random high and random low trial types in RT (RT for random low minus RT for random high trials). The learning and memory scores of sequence learning were calculated as the difference between pattern and random high trial types in RT (RT for random high minus RT for pattern trials; Fig. 4.1C). Higher scores indicate larger statistical or sequence learning/memory.

To assess learning and the retention of knowledge, repeated-measures ANOVAs and paired-samples t-tests were conducted on standardized RT data, separately for statistical and sequence learning. The Greenhouse-Geisser epsilon (ε) correction was used when necessary. Original *df* values and corrected *p* values (if applicable) are reported with partial eta-squared (η^2_p) as a measure of effect size. For correlation analyses, in case of normal distribution, Pearson's correlation was employed. When the assumption of normal distribution was violated, Spearman correlation was used for frequentist statistics and Kendall's Tau-b correlation was used for Bayesian statistics.

In conjunction with the frequentist analyses, we performed Bayesian paired-samples t-tests and calculated the Bayes Factor (BF) for the relevant comparisons as well. The BF is an excellent tool that helps to conclude whether the collected data supports the null-hypothesis (H_0) or the alternative hypothesis (H_1) (Wagenmakers et al., 2011). BFs can be particularly relevant in consolidation studies where memory retention is reflected by evidence supporting the H_0 rather than H_1 (Dienes, 2014). In this case, H_0 states the lack of difference between the mean of the memory scores before and after the offline period, while H_1 means the mean that the memory scores differ. Here, we report BF_{01} values, which were calculated using the JASP software (version 0.11.1.0., JASP, 2019). According to Wagenmakers et al. (2011) BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 suggest substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis.

3. Results

3.1. Prerequisite of memory consolidation

To assess memory consolidation, significant learning has to occur preceding the offline period. Therefore, as a first step, we conducted repeated-measures ANOVAs on the Learning Phase to confirm that significant learning has occurred concerning both statistical and sequence learning. ANOVAs were conducted on standardized RT data, separately for statistical and sequence learning.

Statistical learning during the Learning Phase was tested with a two-way repeated-measures ANOVA on RT with PROBABILITY (random high vs random low) and EPOCH (1-4) as within-subject factors. The ANOVA showed significant statistical

learning (main effect of PROBABILITY, $F(1, 69) = 128.65$, $p < .001$, $\eta^2_p = .65$; Fig. 4.2A), participants showed faster responses to random high ($M = 0.93$) compared to random low trials ($M = 0.98$). RTs gradually decreased as the task progressed, irrespective of trial types (main effect of EPOCH, $F(3, 207) = 50.29$, $p < .001$, $\eta^2_p = .42$). The RT difference between random high and random low trials did not change throughout the task (non-significant PROBABILITY × EPOCH interaction, $F(3, 207) = 2.25$, $p = .084$).

To test **sequence learning** during the Learning Phase, a similar two-way repeated-measures ANOVA on RT with ORDER (pattern vs random high) and EPOCH (1-4) as within-subject factors were conducted. The ANOVA confirmed significant sequence learning (main effect of ORDER, $F(1, 69) = 6.09$, $p = .02$, $\eta^2_p = .08$; Fig. 4.2B). Pairwise comparisons showed faster RTs to pattern ($M = 0.90$) than to random high trials ($M = 0.93$). RTs gradually decreased as the task progressed, irrespective of trial types (main effect of EPOCH, $F(3, 207) = 86.85$, $p < .001$, $\eta^2_p = .56$). Moreover, participants were increasingly faster on pattern trials than on random high trials as the task progressed (revealed by the significant ORDER × EPOCH interaction, $F(3, 207) = 4.43$, $p = .02$, $\eta^2_p = .06$).

Furthermore, to investigate whether individual differences influence the learning on the task, we correlated statistical and sequence learning scores with working memory capacity, with percentage of perseverative errors on the WCST task, with socioeconomic status, and with total problem score on the SDQ. To control for multiple comparisons, we employed False Discovery Rate correction. None of the correlations were significant (all *p*s > .064). We also rerun the ANOVAs on the sample without left-handed participants to control for handedness. The results were identical to the ones on the whole sample.

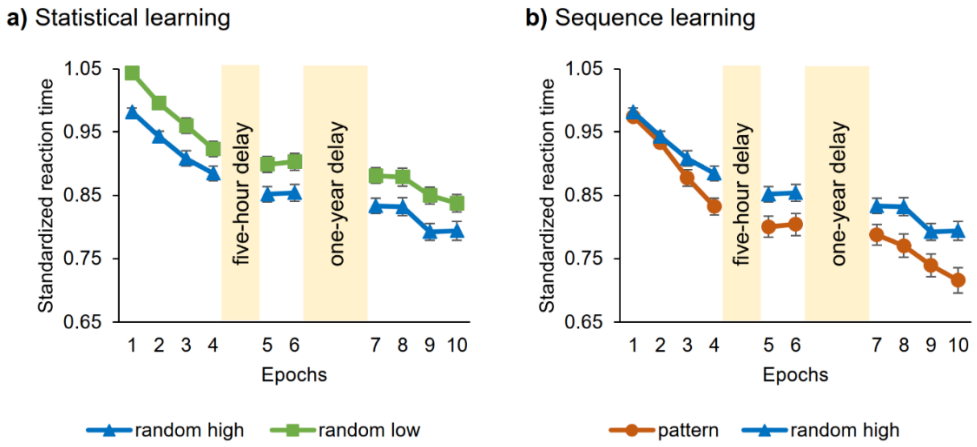

Figure 4.2. Temporal dynamics of (A) statistical and (B) sequence learning across epochs and sessions. Standardized RT values as a function of the epoch (1-10) and trial types (random high vs random low for statistical learning and pattern vs random high for sequence learning) are

presented. Blue lines with triangle symbols indicate standardized RT values on the random high trials, green lines with square symbols indicate standardized RT values on the random low trials and orange lines with circle symbols indicate standardized RT values on the pattern trials. (A) Statistical learning is quantified by the gap between blue and green lines and (B) sequence learning is quantified by the gap between orange and blue lines. In both cases, greater gap between the lines represents better learning. Error bars denote standard error of mean.

3.2. Do children retain regularities after a one-year offline period?

To test one-year consolidation of **statistical knowledge**, we conducted a two-way repeated-measures ANOVA on RT with PROBABILITY (random high vs random low) and EPOCH (6 vs. 7) as within-subject factors. Overall, irrespective of epochs, participants were faster on random high (M = 0.84) than on random low trials (M = 0.89) (main effect of PROBABILITY, $F(1, 69) = 159.11$, $p < .001$, $\eta^2_p = .70$). Average RTs (i.e., RTs irrespective of trial types) differed in the two epochs (main effect of EPOCH, $F(1, 69) = 3.92$, $p = .05$, $\eta^2_p = .05$), participants showed faster RTs in the 7th epoch (M = 0.86) compared to the 6th epoch (M = 0.88). Crucially, the ANOVA revealed evidence for retained statistical memory after the one-year delay (non-significant PROBABILITY × EPOCH interaction, $F(1, 69) = 0.03$, $p = .86$, $BF_{01} = 7.50$; Fig. 4.3A), with similar memory scores in the 6th (M = 0.049) and in the 7th (M = 0.048) epochs. Furthermore, as the delay has some variability in terms of weeks (M_{delay} = 53.08 weeks, SD_{delay} = 2.39 weeks, between 47.95 and 60.24 weeks), we examined whether it has any relation to the long-term memory performance. First, we calculated an offline change score for statistical knowledge by subtracting the standardized memory score in Epoch 6 from the standardized memory score in Epoch 7. This way, negative scores indicate forgetting and positive scores indicate offline learning. Offline change score did not show correlation with the length of the long-term delay ($rs(68) = .071$, $p = .558$; $BF_{01} = 5.115$).

To investigate one-year consolidation of **serial-order knowledge**, we also run a two-way repeated-measures ANOVA on RT with ORDER (pattern vs random high) and EPOCH (6 vs. 7) as within-subject factors. Overall, participants showed faster RTs on pattern (M = 0.80) than on random high trials (M = 0.84) (main effect of ORDER, $F(1, 69) = 5.88$, $p = .018$, $\eta^2_p = .078$). Average RTs were similar in the two epochs (main effect of EPOCH, $F(1, 69) = 3.33$, $p = .073$). Importantly, the ANOVA revealed retained serial-order knowledge (non-significant ORDER × EPOCH interaction, $F(1, 69) = 0.18$, $p = .67$, $BF_{01} = 6.97$; Fig. 4.3B), memory scores were similar in the Testing and Retesting Phases (6th epoch: M = 0.05, 7th epoch: M = 0.045). Similarly to statistical knowledge, we also correlated the offline change score of serial-order knowledge and the length of the

long-term delay. One participant had to be excluded from the analysis as they showed extremely low offline change score according to Tukey (1977) criterion (more than 1.5 times the interquartile range). Offline change scores did not correlate with the length of the delay ($rs(67) = -.085$, $p = .485$; $BF_{01} = 5.011$).

Moreover, similarly for the learning scores, to investigate whether individual differences influence the consolidation of statistical or serial-order knowledge, we correlated the offline change scores with working memory capacity, with percentage of perseverative errors on the WCST task, with socioeconomic status, and with total problem score on the SDQ. To control for multiple comparisons, we employed False Discovery Rate correction. None of the correlations reached significance (all *p*s > .277). We also rerun the ANOVAs on the sample without left-handed participants to control for handedness. The results were identical to the ones on the whole sample.

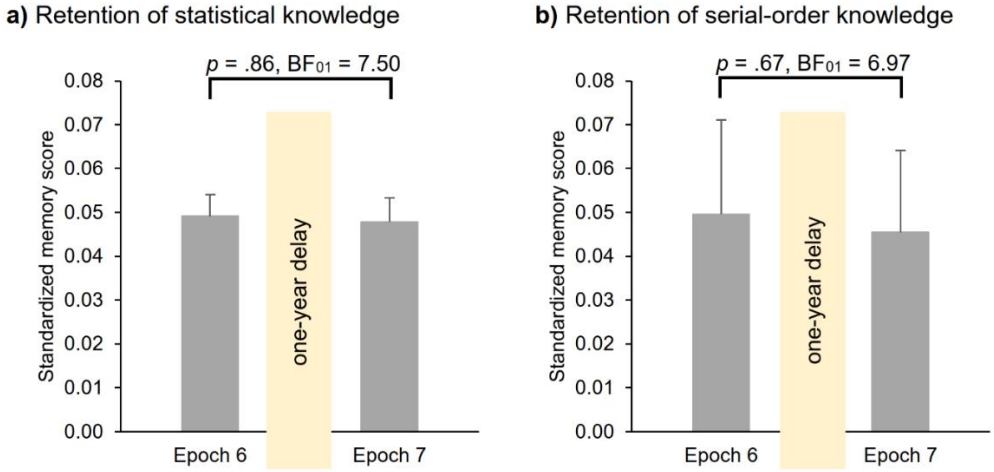

Figure 4.3. Retention of (A) statistical and (B) serial-order knowledge. Memory scores measured by standardized RT values for the last epoch of the Testing Phase (Epoch 6) and the first epoch of the Retesting Phase (Epoch 7). Error bars denote the standard error of mean.

3.3. Does age affect the one-year retention of statistical and serial-order regularities?

To check the possible association between age and retention, we conducted Pearson's correlation between the offline change scores and age. Regarding statistical knowledge, offline change scores did not show correlation with age ($r(68) = .06$, $p = .62$, $BF_{01} = 5.92$). Concerning serial-order knowledge, one participant had to be excluded from the analysis on RT data as they showed extremely low offline change score according to Tukey (1977) criterion (more than 3 times the interquartile range). The correlation between the offline change score of serial-order knowledge represented by RT values and age was also not significant ($r(67) = -.06$, $p = .62$, $BF_{01} = 5.91$).

4. Discussion

The present study aimed to investigate the one-year consolidation of probability-based and serial order-based regularities in children aged 9 to 15 years. We have shown retained knowledge of both information after the one-year offline period; participants successfully learnt and stabilized the regularities, and the acquired knowledge was resistant to forgetting over a long period of time. Additionally, successful retention was irrespective of age as we have not found association between the retention of probability-based or serial order-based regularities and age. Our results are supported by Bayesian statistics as well, further strengthening the evidence for successful one-year retention.

Both statistical and serial-order knowledge has been successfully retained after the one-year offline period, which is in accordance with previous adult studies (Kóbor et al., 2017; Romano et al., 2010). However, Romano et al. (2010) and Kóbor et al. (2017) employed the original, uncued version of the ASRT task in which probability-based and serial order-based regularities are intermixed. Here, we went beyond these studies by employing the cued version of the ASRT task which is designed to disentangle these two types of regularities. This version dissects probability-based and serial order-based regularities by marking pattern and random elements with different visual stimuli. This modification results in the possibility of measuring the acquisition of probability-based and serial order-based regularities (i.e., statistical and sequence learning, respectively) within the same experimental design and within the same time window (Nemeth, Janacsek, & Fiser, 2013). It is important to note that the dissection of probability-based and serial order-based regularities in the ASRT task is possible even in the original, uncued version of the task (i.e., where pattern and random stimuli are not visually distinguishable), however, excessive training (i.e., four-day-long practice) is needed to reach that aim (Howard, Howard, Japikse, DiYanni, et al., 2004; Howard, & Howard, 1997). By implementing the cued version of the ASRT task in our study design, we could simultaneously measure the encoding and consolidation of probability-based and serial order-based information, enabling us to compare the consolidation of these two processes. Although several empirical studies have shown the differences between statistical and sequence learning in encoding, to the best of our knowledge, only two studies have investigated the consolidation of these processes within the same experimental design. Simor et al. (2019) have found retained knowledge after a one-hour offline period with no difference on the behavioral level between statistical and serial-order knowledge.

However, on the neural level, they discovered that slow frequency oscillations (high delta and theta power) during sleep predicted further improvements in sequence learning, while changes in statistical learning were not associated with spectral EEG power measures. Zavecz et al. (2020) also explored the brain activity underlying the consolidation of probability-based and serial order-based information. Although the consolidation of probability-based and serial order-based information was comparable on the behavioral level showing successful retention of both types of knowledge, differences emerged on the neural level. Consolidation of statistical knowledge was in relation with learning-induced changes in delta-frequency connectivity between local, short-range connections, while consolidation of serial-order knowledge was associated with learning-induced changes in alpha frequency connectivity over long-range centro-parietal networks. Taken together, the present study corroborates these findings as we also showed retention of statistical and serial-order knowledge after a one-year offline period on the behavioral level. Further studies are warranted to examine brain activity underlying long-term consolidation of probability-based and serial order-based information.

The findings of retained statistical and serial-order knowledge in children after a long period of time extends previous studies showing retention or even offline learning over the short or medium term. In more detail, Fischer et al. (2007) showed retained statistical knowledge after an 11-hour offline period spent awake; whereas Desmottes et al. (2017) investigated sequence-specific learning and found offline learning both after 24-hour and one-week delay, and Hedenius, et al. [32] found retained serial-order knowledge following a 24-hour delay. Four studies (Hedenius et al., 2020; Hedenius et al., 2013; Hedenius et al., 2011; Takács et al., 2018)[7] employed the uncued ASRT task (intermixing probability-based and serial order-based regularities). They found retained knowledge following a 16-hour (Nemeth, Janacsek, Balogh, et al., 2010; Takács et al., 2018) and three-day (Hedenius et al., 2011) offline period, and offline learning following a 24-hour delay (Hedenius et al., 2020). Our results on RT data are consistent with these studies. Although here we focused on RT data, analyses on accuracy data also yielded similar results (see Supplementary Material). Altogether, our results corroborate and extend the previous ones with showing successful retention after a one-year long offline period.

Although unveiling lifespan differences in the consolidation of statistical and serial-order knowledge was not the goal of the present study, it is worth noting that our results are in line with the findings of Romano et al. (2010) and Kóbor et al. (2017),

showing successful one-year retention in adults. The development and lifespan trajectory of the *acquisition* of probability-based and serial order-based information underwent thorough investigation (e.g., Janacsek et al., 2012; Lukács & Kemény, 2015; Nemeth, Janacsek, & Fiser, 2013; Reber, 1993); however, no consensus emerged whether learning is age-dependent or not (for a review, see Zwart et al., 2019). Nemeth, Janacsek, and Fiser (2013) examined the acquisition of probability-based and serial order-based regularities employing both the uncued and the cued ASRT task in a population of neurotypicals between the age of 11 and 39. In the uncued condition, 11-13-year-old children showed higher extent of statistical learning compared to the other age groups. This falls in line with the 'less is more' model, which proposes age-dependent learning of regularities with a peak performance during childhood, up until the age of 12 (Janacsek et al., 2012; Juhasz et al., 2019). In contrast, statistical learning was age-invariant in the cued condition. Sequence learning was similar in all age groups in both the cued and uncued conditions (Nemeth, Janacsek, & Fiser, 2013), suggesting that the acquisition of serial order-based information is comparable from childhood to adulthood. Here, we went beyond the study of Nemeth, Janacsek, and Fiser (2013) by investigating the *consolidation* of probability-based and serial order-based regularities. Importantly, not only learning but consolidation of statistical and serial-order knowledge could also differ during the lifespan (e.g., Adi-Japha, Badir, Dorfberger, & Karni, 2014; Fischer et al., 2007). Fischer et al. (2007) showed age-dependent consolidation of statistical knowledge in the case of sleep-dependent consolidation. Adults benefited from sleep and showed better consolidation of statistical knowledge after sleep than wakefulness, while the exact opposite picture emerged in children (however, for the confounding effect of pre-sleep level performance, see Wilhelm, Metzkow-Mészàros, Knapp, & Born, 2012). As for the consolidation of serial-order knowledge, the results of Adi-Japha et al. (2014) suggests a more nuanced picture: memory performance in childhood and adulthood on the behavioral level appeared similar, showing retention of knowledge in both age groups, however, children seemed to be less susceptible to subsequent interference than adults. In the present study, we found long-lasting representation of statistical and serial-order knowledge, similarly to the studies of Romano et al. (2010) and Kóbor et al. (2017) that showed retained statistical knowledge following a one-year offline period in adults. Thus, our study offers indirect evidence of comparable consolidation of probability-based and serial order-based information in childhood and adulthood, supporting developmental invariance in consolidation. The lack of association between retention and age in our sample also

promotes the developmental invariance model. Nevertheless, as we did not directly contrast the performance of adult and children populations, further studies are warranted to examine the long-term memory performance of statistical and serial-order knowledge in adults and children within the same experimental design.

It is also worth looking at our results from a broader perspective of memory consolidation, namely the distinction between procedural and declarative processes. Statistical and serial-order regularities have been proposed to be two aspects of procedural memory, i.e., the system underlying the acquisition of skills and habits (Howard, & Howard, 1997; Nemeth, Janacsek, & Fiser, 2013). Hence, our results, together with previous studies on adults, suggest that consolidation of procedural knowledge is age-invariant and comparable from childhood to adulthood, both after short-term and long-term delay. The developmental differences of declarative memory, i.e., the system underlying the learning and remembering of facts and events, has also been investigated across the lifespan. While declarative memory abilities have been extensively shown to improve across childhood and adolescence, particularly memory for contextual details (e.g., Bouyeure & Noulhiane, 2020; Gulya et al., 2002; Picard, Cousin, Guillery-Girard, Eustache, & Piolino, 2012), the developmental differences of declarative memory consolidation using relatively long offline periods (i.e., more than 24 hours) have been tested only in a handful of studies. For example, Henderson, Weighall, Brown, and Gareth Gaskell (2012) showed that children have retained knowledge of object' locations following a one-week delay. Relatedly, in school-aged children, cued recall of previously learnt novel words was maintained (Smith et al., 2018) or even improved (Henderson et al., 2012) after a one-week offline period. Recognition of novel words was also maintained after a one-week delay (Henderson et al., 2012). Similarly, in neurotypical adults, Gaskell and Dumay (2003) showed retained explicit recognition of priorly acquired novel words and Dumay, Gaskell, and Feng (2004) found increased free recall of novel words after a one-week offline period. In sum, similarly to procedural memory, long-term consolidation of declarative memory (one-week delay in these examples) seems to be comparable between school-aged children and adults, at least considering the learning and recalling of novel words. Note that, to the best of our knowledge, no one-year consolidation has been tested for this (or other) aspect of declarative memory in children that would allow greater comparability. Future studies are warranted to directly compare the developmental trajectory of the long-term consolidation of procedural and declarative memory within the same groups using a range of offline delays.

In the present study, we took into consideration several possible confounds of consolidation. Participants completed the task three times: (1) in the Learning Phase, (2) in the Testing Phase five hours later on the same day and (3) in the Retesting Phase one year later, with no practice during the offline periods. Employing this study design, we controlled for the following possible confounds. First, by implementing a Testing Phase in the design, we controlled for the short-term (five-hour) consolidation of information. Second, as participants were unaware of the fact that they will be tested with the same task later, any confounding effects of explicit strategy during acquisition or consolidation were minimized. Lastly, during the offline periods, there was no practice, which could have led to the reactivation of the acquired knowledge. Moreover, we took into account the possible confounding effect of individual differences on consolidation. We correlated consolidation performance with working memory capacity, executive functions, socioeconomic status, behavioral and emotional problems, and examined the role of handedness as well. We did not find any relation between these factors and consolidation performance; therefore, it is highly unlikely that individual differences confounded our results.

Taken together, the present study demonstrated that the representation of statistical and serial order-based regularities remains stable over a long period of time in neurotypical children and can be successfully retained after a one-year offline period. We showed that the knowledge of statistical and serial order-based regularities is robust and resistant to forgetting over a one-year offline period, with no difference between the two aspects of learning. Our study also offers indirect evidence for the developmental invariance of consolidation of statistical and serial-order knowledge.

V. Study 4: Lifespan developmental invariance in memory consolidation: evidence from procedural memory

Publication:

Tóth-Fáber, E., Nemeth, D. & Janacsek, K. (2023). Lifespan developmental invariance in memory consolidation: evidence from procedural memory. *PNAS Nexus*, pgad037. https://doi.org/10.1093/pnasnexus/pgad037

Abstract

Characterizing ontogenetic changes across the lifespan is a crucial tool in understanding neurocognitive functions. While age-related changes in learning and memory functions have been extensively characterized in the past decades, the lifespan trajectory of memory consolidation, a critical function that supports the stabilization and long-term retention of memories, is still poorly understood. Here we focus on this fundamental cognitive function and probe the consolidation of procedural memories that underlie cognitive, motor, and social skills and automatic behaviors. We used a lifespan approach: 255 participants aged between 7 and 76 performed a well-established procedural memory task in the same experimental design across the whole sample. This task enabled us to disentangle two critical processes in the procedural domain: statistical learning and general skill learning. The former is the ability to extract and learn predictable patterns of the environment, while the latter captures a general speed-up as learning progresses due to improved visuomotor coordination and other cognitive processes, independent of acquisition of the predictable patterns. To measure the consolidation of statistical and general skill knowledge, the task was administered in two sessions with a 24-hour delay between them. Here, we report successful retention of statistical knowledge with no differences across age groups. For general skill knowledge, offline improvement was observed over the delay period, and the degree of this improvement was also comparable across the age groups. Overall, our findings reveal age invariance in these two key aspects of procedural memory consolidation across the human lifespan.

Significance statement

Consolidation is a critical function responsible for the stabilization and long-term retention of memories. Here, we tested the consolidation of procedural memories, which underlie skills and automatic behaviors, using a lifespan approach. In contrast to the age-variant lifespan trajectory of procedural learning, our results revealed age-invariant procedural memory consolidation across the lifespan. Thus, procedural learning and consolidation seem to follow distinct developmental curves in neurotypical individuals. These findings suggest at least partially different neural underpinnings of learning versus consolidation and will likely stimulate future neuroimaging research and theory development of memory.

1. Introduction

Identifying age-related changes in cognitive functions across the human lifespan is a crucial step in understanding brain development and developing more efficient diagnostic tools and interventions for developmental delays and decline in old age. Substantial research has focused on characterizing how cognitive functions change across the lifespan. The largest body of evidence comes from studies comparing the cognitive performance of typically two to four age groups. There are comparably fewer large-scale, cross-sectional or longitudinal studies that track performance from childhood to older adulthood (i.e., across the lifespan) using the same task. These large-scale studies, however, are essential for a better understanding of cognitive changes across the lifespan as they control for a range experimental and analytical factors that cannot be controlled when lifespan trajectories are inferred based on a diverse set of individual studies. Combined evidence from these different study approaches reveals markedly distinct lifespan trajectories depending on the cognitive function of interest. For example, aspects of executive functions (Cepeda, Kramer, & Gonzalez de Sather, 2001; Zelazo, Craik, & Booth, 2004), working memory (Alloway & Alloway, 2013; Borella, Carretti, & De Beni, 2008; Conklin, Luciana, Hooper, & Yarger, 2007), autobiographical memory (Rathbone, Moulin, & Conway, 2008) and episodic memory (Bauer et al., 2013; Dikmen et al., 2014) have been shown to follow an inverted U-shape trajectory, with continuous maturation during childhood, a peak performance in young adulthood, and a deterioration in older adulthood. In contrast, language acquisition, general skill learning, and statistical learning—the ability to extract and learn predictable patterns of the environment—seem to peak during childhood, followed by a decline in adulthood (Janacsek et al., 2012; Johnson & Newport, 1989; Juhasz et al., 2019). Furthermore, certain cognitive functions might remain intact later in life as well, such as automatic processes of memory retrieval (Ikier, Yang, & Hasher, 2008). While our knowledge on the age-related changes in learning and memory functions has greatly expanded in the past decades, the lifespan trajectory of memory consolidation—a critical function that is responsible for the stabilization and long-term retention of memories—is still poorly understood. Here we focus on this fundamental cognitive function and probe the consolidation of memories acquired via statistical learning using a lifespan approach.

Statistical learning is a crucial aspect of life from infancy to old age as it enables us to extract complex probabilistic regularities embedded in the environment, allowing

us to adapt to our surroundings throughout the human lifespan (Armstrong, Frost, & Christiansen, 2017; Aslin, 2017; Fiser & Aslin, 2002; Saffran, Aslin, & Newport, 1996; Turk-Browne, Scholl, Johnson, & Chun, 2010). Through extensive practice, statistical learning contributes to the acquisition of automatic behaviors, such as skills and habits, which are rooted in procedural memory (Hallgato et al., 2013; Kaufman et al., 2010; Romano et al., 2010; Ullman, 2004). The developmental trajectory of statistical learning has been described with three different models (Zwart et al., 2019). The age-invariant model suggests no developmental changes across the lifespan (Reber, 1993), based on studies showing comparable learning performance in children and adults (e.g., Meulemans et al., 1998) and based on results showing that statistical learning is related to brain regions that mature early, such as the striatum (Reber, 1993). The other two models propose that statistical learning varies as a function of age. The inverted U-shaped model suggests a gradual improvement over childhood and adolescence, with the best performance in young adulthood and a decline with aging (Lukács & Kemény, 2015). This model is supported by results finding better learning performance in young adulthood than in childhood and old adulthood (e.g., Thomas et al., 2004). Involving a large sample of participants from childhood to old adulthood, a study (Lukács & Kemény, 2015) found evidence for the inverted U-shaped model examining participants between 7 and 87 years of age. The third model, which can be referred to as 'competition model', argues for better statistical learning in childhood (under the age of 12), less effective learning in adolescence and adulthood and a decline in old adulthood (Janacsek et al., 2012; Juhasz et al., 2019). In detail, Janacsek et al. (2012) differentiate between the detection of raw probabilities and the usage of internal models. They argue that due to the yet underdeveloped internal models, children are more sensitive to raw statistical probabilities of the environment, which translates to better statistical learning performance. The development of internal models in adolescence and adulthood then leads to less reliance on raw statistical probabilities as more complex interpretations of the observed probabilities emerge. The decline in old adulthood can be explained by reduced sensitivity to raw statistical probabilities, increased rigidness of internal models, and/or a weaker connection between these two systems. Employing a lifespan approach, a study (Janacsek et al., 2012) investigated participants from the age of 4 to 85 years, showing better statistical learning under the age of 12. Moreover, Nemeth, Janacsek, and Fiser (2013) contrasted the performance of five age groups from 11 to 39 years. They showed better statistical learning in the 11-13-year-old group compared to the other age

groups, while statistical learning was similar from the age of 14 to 39. A recent study by Juhasz et al. (2019) examined statistical learning from the age of 7 to 85 years involving the same pool of participants as the present study. Notably, Juhasz et al. (2019) focused on the trajectory of statistical learning and general skill learning (see below for details), whereas the present study focuses on the 24-hour consolidation of such knowledge.

Importantly, statistical learning does not occur only during practice but also between the practices, in the so-called offline periods. Via consolidation, the initially fragile and unstable memory representations are converted into a more stable form, ensuring that they are preserved and can be retrieved later (Walker, 2005). Successful consolidation can be reflected by retention (i.e., no forgetting, similar performance at the end of learning and during subsequent testing) or even by offline gains (i.e., offline learning, better performance during testing than at the end of learning) (Robertson et al., 2004). The consolidation of knowledge acquired via statistical learning has been tested across different time delays (e.g., from hours to days or even a year) between learning and testing, but all studies have focused on one age group at a time or contrasted performance of a couple of age groups (e.g., children vs. adults; young vs. older adults) only. The present study aims to go beyond previous research by examining consolidation of statistical knowledge across the lifespan, in a sample of participants aged between 7 and 76 years.

Despite the ample investigation on the lifespan trajectory of statistical learning, the *consolidation* of such knowledge did not receive much attention. To the best of our knowledge, no models were proposed for the lifespan trajectory of the consolidation of statistical knowledge. Considering the proposed trajectories of statistical learning, different developmental curves can be proposed for the consolidation of statistical knowledge. As described above, two age-variant trajectories have been proposed for the development of statistical learning (Janacsek et al., 2012; Lukács & Kemény, 2015). It raises the question whether we can expect that the consolidation of such knowledge will also follow an age-variant trajectory. In atypical development, it has been demonstrated that learning and consolidation can show dissociation: Enhanced learning and intact consolidation has been shown in Tourette syndrome (Takács et al., 2018), whereas impaired consolidation has been shown to accompany intact learning in developmental dyslexia (Hedenius et al., 2021). However, it is still an open question whether learning and consolidation show a dissociation in neurotypical populations, especially across development and aging.

As described above, research on the consolidation of statistical knowledge has focused on one age group at a time or contrasted performance in a couple of age groups only. Most studies have suggested that children and adolescents can successfully retain the acquired knowledge following delays ranging from hours to one-year (Hedenius et al., 2013; Hedenius et al., 2011; Nemeth, Janacsek, Balogh, et al., 2010; Takács et al., 2018; Tóth-Fáber, Janacsek, et al., 2021), while others have found offline learning (i.e., improved performance) in a group of children and adolescents after a 24-hour delay (Hedenius et al., 2020). Smalle, Page, Duyck, Edwards, and Szmalec (2018) investigated a related process, that is, Hebb learning in a longitudinal design with 8-9-year-old children and adults and tested the retention of sequences over a four-hour, one-week and one-year offline delay. The results showed better consolidation in children compared to adults across all offline periods. In young and middle aged adults, knowledge of statistical regularities seems to be successfully retained, irrespective of the length of delay (e.g., Arciuli & Simpson, 2012; Horvath et al., 2020; Kim et al., 2009; Kóbor et al., 2017; Meier & Cock, 2014; Nemeth, Janacsek, Király, et al., 2013; Romano et al., 2010; Simor et al., 2019; Song et al., 2007b). In contrast, the handful of studies focusing on older adults have revealed mixed results: Some studies have suggested retention of the acquired knowledge (Romano et al., 2010), whereas others have indicated a decline over the delay period (Nemeth & Janacsek, 2011). Overall, based on these studies, no firm conclusions could be drawn on the consolidation of statistical knowledge across the lifespan, although retention (that is, no performance change) seems to be the most plausible outcome for most age groups, which would support an age-invariant model of the consolidation of statistical knowledge.

Since statistical learning requires repeated exposure to the same regularities (Conway, 2020), during this period of repeated exposure (i.e., in the learning phase), other learning processes are also engaged that could confound the measures and interpretation of statistical learning as well as its consolidation. One such learning process is called general skill learning, which refers to the faster processing of and responding to stimuli and improved visuomotor coordination as a result of practice, independent of the regularities embedded in the stimulus stream (Hallgato et al., 2013; Juhasz et al., 2019). In the present study, we use a task design that enables us to tease apart consolidation processes specific to statistical knowledge by contrasting it to the consolidation of general skill knowledge. Unveiling the lifespan trajectory of the consolidation of statistical and general skill knowledge using a carefully controlled, identical design across age groups

from childhood to older adulthood can significantly improve our understanding of how the consolidation of different types of knowledge changes across development and aging and can shed light on the age-related changes in brain plasticity supporting these functions.

With regards to the consolidation of general skill knowledge, offline improvement has been shown both in children and adults, that is, participants usually exhibited faster average reaction times after an offline period (Csábi et al., 2013; Csabi et al., 2015; Hedenius et al., 2020; Hedenius et al., 2013; Nemeth, Janacsek, Balogh, et al., 2010; Nemeth, Janacsek, Király, et al., 2013). Nevertheless, whether the extent of offline improvement differs across development from childhood to adulthood remains unclear. In older adults, the results on the consolidation of general skill knowledge are largely mixed. Elderly participants demonstrated offline gains over a 12-hour delay (Nemeth & Janacsek, 2011; Nemeth, Janacsek, Londe, et al., 2010) , but the gain was smaller than in young adults (Nemeth, Janacsek, Londe, et al., 2010). Moreover, Nemeth and Janacsek (2011) did not find evidence for improvement following 24-hour and one-week delays in older adults, while young adults showed significant improvements following both delay periods. Retention but no offline improvement of general skill knowledge has been found over a one-year delay as well, with no differences between young and older adults (Romano et al., 2010). Thus, while offline improvement of general skill knowledge may be expected in some cases (e.g., for shorter delays), overall, no conclusive pattern across studies could be revealed, especially for potential differences in the extent of this improvement from childhood to adulthood. Nonetheless, based on the previous studies, it is reasonable to expect at least some age-variance for the consolidation of general skill knowledge.

To the best of our knowledge, no study has tested consolidation of statistical and/or general skill knowledge with the same experimental design across the lifespan so far. The present study fills this gap using a learning task that enables us to tease apart consolidation processes specific to statistical knowledge vs. general skill knowledge in a large sample of participants aged between 7 and 76 years. By employing the same experimental design across the whole sample, our study can unveil the lifespan trajectory of the consolidation of statistical and general skill knowledge: Crucially, it can provide clear evidence for potential differences in consolidation across age groups from childhood to older adulthood as well as across knowledge types. Based on the previous empirical findings, for the lifespan trajectory of consolidation of statistical knowledge, an age-

invariant trajectory can be proposed, whereas the consolidation of general skill knowledge might follow an age-variant trajectory. The findings of the present study can greatly improve our understanding of the consolidation of different types of knowledge across the human lifespan and can shed light on the age-related changes in brain plasticity supporting these functions. Unveiling the lifespan trajectory of statistical as well as general skill knowledge can also help develop a theoretical model for these processes.

2. Methods

2.1. Participants

Two hundred-seventy participants took part in the present study. They were assigned to nine age groups (n = 30 in each group). Fourteen participants were excluded based on outlier (above 3 *SD*s) performance in average response times or accuracy during the whole experiment compared to their respective age group. The developmental trajectory of statistical learning and general skill improvements in the Learning Phase of this sample are reported in Juhasz et al. (2019). The present study focuses on the consolidation of statistical knowledge and general skills; these results were not reported elsewhere. For consistent age distribution, we decided to exclude one (85-year-old) participant from the oldest age group due to being outlier in terms of age. Hence, the final sample of the present study consisted of 255 participants aged between 7 and 76 years. Mean and standard deviation for age and gender ratio for all age groups are presented in Table 5.1. Caregivers of underage participants completed a parental questionnaire and adults completed a self-report questionnaire regarding health-related questions. All participants had normal or corrected-to-normal vision and none of the participants had any neurological, psychiatric, or neurodevelopmental disorder. Adult participants gave informed written consent, whereas caregivers of underage participants provided informed written consent and children and adolescents provided verbal consent to participate in the study before enrollment. Participants received no financial compensation for participation. All experimental procedures were approved by University Research Ethics Committee, and was conducted in accordance with the Declaration of Helsinki.

Table 5.1. Demographic data (mean and standard deviation for age, and gender ratio) for all age groups.

Group	Age	Gender
7-8-year-old (n=26)	7.92 (0.27)	13 M / 13 F
9-10-year-old (n=28)	9.79 (0.42)	13 M / 15 F
11-13-year-old (n=30)	12.10 (0.61)	13 M / 17 F
14-15-year-old (n=30)	14.55 (0.57)	13 M / 17 F
16-17-year-old (n=30)	16.56 (0.54)	13 M / 17 F
18-29-year-old (n=30)	21.64 (2.93)	12 M / 18 F
30-44-year-old (n=30)	36.67 (3.81)	12 M / 18 F
45-60-year-old (n=26)	51.65 (4.46)	6 M / 20 F
61-76-year-old (n=25)	65.28 (4.47)	5 M / 20 F

2.2. Task

The Alternating Serial Reaction Time (ASRT) task was used to assess statistical learning and consolidation (Howard, & Howard, 1997; Nemeth, Janacsek, Londe, et al., 2010). In this task, four horizontally arranged empty circles are presented on the screen and a stimulus (a dog's head) appeared in one of the circles (Nemeth, Janacsek, Polner, & Kovacs, 2013). Participants were instructed to press a corresponding key (Z, C, B, or M on a QWERTY keyboard) as quickly and accurately as they could when the stimulus occurred using their index and middle fingers. After the correct response of the participant, the next stimulus appeared 120 ms later. Unbeknownst to the participants, the presentation of stimuli followed an eight-element sequence, within which pattern (P) and random (r) trials alternated with each other (e.g., 2-*r*-4-*r*-3-*r*-1-*r*; where numbers indicate the four locations on the screen from left to right, and *r* denote a randomly chosen location out of the four possible ones; see Fig. 1).

Due to this alternating sequence, some runs of three consecutive trials (triplets) were more probable than others. In the example sequence 2-*r*-4-*r*-3-*r*-1-*r*, triplets 2-X-4, 4-X-3, 3-X-1, and 1-X-2 (where X indicates the middle element of the triplet) occurred with a higher probability because they were presented in every sequence repetition (P-*r*-P) and could also be formed by chance (*r*-P-*r*, see Fig. 5.1B). Note that here, we use X to indicate the middle element of the triplet because, for example, 4-X-3 (e.g., 4-2-3 in Fig. 5.1B) can appear both as a P-*r*-P structure (where the first and last element of the triplet belong to the predetermined pattern) and as a *r*-P-*r* structure (where the first and last elements are random, and the middle element is part of the predetermined pattern). In contrast, triplets 2-X-1 and 3-X-2 occurred with a lower probability since they could only

be formed by chance (that is, their structure could only be *r*-P-*r*). The former triplet types are referred to as high-probability triplets and the latter ones as low-probability triplets. Overall in the task, high-probability triplets were five times more probable than the low-probability ones (Kóbor et al., 2017; Nemeth, Janacsek, & Fiser, 2013). Note that triplets were identified using a moving window throughout the stimulus stream. Thus, each trial was categorized as the third element of a high- or a low-probability triplet, and this categorization was used in our analyses; the same trial then served as the middle and the first element for the categorization of the following triplets.

The ASRT task enables us to separate statistical learning from general skill improvements. Statistical learning is defined as faster and more accurate responses to high-probability elements than to low-probability ones (Howard, & Howard, 1997). In contrast, general skill improvements refer to average speed-up and changes in accuracy that are independent of the probabilities of events. These improvements reflect more efficient visuomotor and motor-motor coordination due to practice (Hallgato et al., 2013; Juhasz et al., 2019).

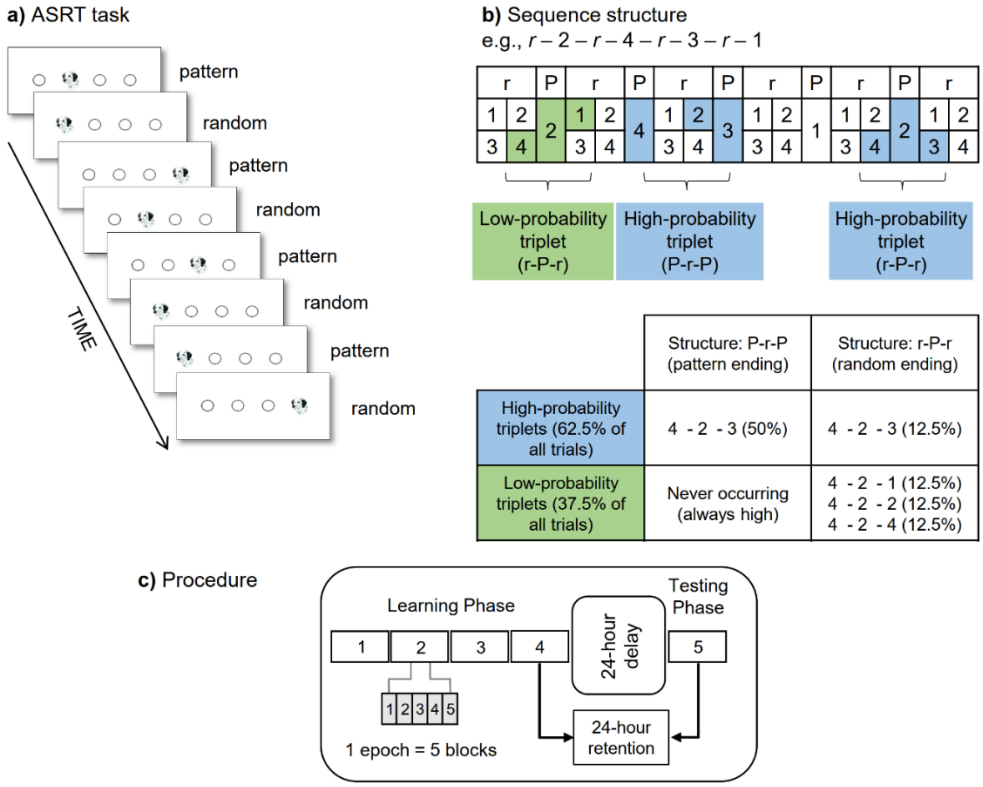

Figure 5.1. The Alternating Serial Reaction Time (ASRT) task. (A) Pattern and random trials were presented in an alternating fashion; the trial types were indistinguishable on the surface level: a picture of a dog's head served as stimuli in all trials. The alternating sequence was coded by the location of stimuli. In pattern trials, the location of stimuli was predetermined, and occurred

in the same order throughout the experiment. In random trials, randomly chosen locations out of the four possible ones were presented. **(B)** An example of the sequence structure. Numbers indicate the predetermined stimulus locations in pattern trials, and *r*s indicate randomly selected locations out of the four possible ones. Due to the alternating sequence, some runs of three consecutive trials (triplets) were more probable than others, referred to as high-probability (green shading) and low-probability triplets (blue shading), respectively. Since high-probability triplets could occur as pattern-ending triplets (50% of all trials) and by chance as random-ending triplets (12.5% of all trials), these triplets constituted 62.5% of all trials. Low-probability triplets constituted the remaining 37.5% of the trials; these were all random-ending triplets. Note that triplets were identified using a moving window throughout the stimulus stream: each trial was categorized as the third element of a high- or a low-probability triplet; the same trial then served as the middle and the first element for the categorization of the following triplets. Fig. 5.1 is adapted from Nemeth, Janacsek, and Fiser (2013) and Zavecz et al. (2020).

2.3. Procedure

The ASRT task was presented in blocks. One block consisted of 85 trials: each block started with five random practice trials followed by the eight-element sequence repeated 10 times. After each block, participants received feedback about their general performance, that is, about their average RTs and accuracy. The ASRT task was administered in two sessions with a 24-hour delay between them. The Learning Phase consisted of 20 blocks. The Testing Phase contained five blocks.

For the alternating sequence, there were 24 permutations of the four possible spatial positions for the predetermined order of pattern trials. However, because of the continuous presentation of the stimuli, for instance, the sequences 2-*r*-1-*r*-3-*r*-4, 1-*r*-3-*r*-4-*r*-2, 3-*r*-4-*r*-2-*r*-1, and 4-*r*-2-*r*-1-*r*-3 were considered identical as they consisted of the same triplets. Consequently, there were six *unique* sequence permutations: 1-*r*-2-*r*-3-*r*-4-*r*, 1-*r*-2-*r*-4-*r*-3-*r*, 1-*r*-3-*r*-2-*r*-4-*r*, 1-*r*-3-*r*-4-*r*-2-*r*, 1-*r*-4-*r*-2-*r*-3-*r*, and 1-*r*-4-*r*-3-*r*-2-*r*. One of these unique permutations was selected for each participant in a pseudorandom manner. For a given participant, the sequence permutation remained the same over the Learning and Testing Phases.

In our study, participants were not informed about the underlying probability structure of the sequence, and they did not even know that they were in a learning situation. Nevertheless, potentially emerged explicit knowledge about the structure was probed by a questionnaire at the end of the Testing Phase (Nemeth, Janacsek, Londe, et al., 2010; Song et al., 2007b). None of the participants reported noticing the sequence in the task. Thus an implicit, non-conscious form of learning was tested (Cleeremans & Dienes, 2008; Reber, 1989; Vékony, Ambrus, Janacsek, & Nemeth, 2021). This is in line with previous studies showing that participants remain unaware of the sequence even after

extended practice, or when more sensitive recognition tests are used to assess explicit knowledge (Howard, Howard, Japikse, DiYanni, et al., 2004; Song et al., 2007b).

2.4. Statistical analysis

Statistical analysis was based on previous studies (Kóbor et al., 2017; Nemeth, Janacsek, Londe, et al., 2010; Romano Bergstrom, Howard, & Howard, 2012); to facilitate data processing, epochs of five blocks were analyzed instead of single blocks (e.g., Blocks 1-5 corresponded to Epoch 1, Blocks 6-10 to Epoch 2, and so on). The Learning Phase consisted of four epochs, while the Testing Phase consisted of one epoch. Similarly to previous studies, two types of low-probability triplets, repetitions (e.g., 222, 333) and trills (e.g., 212, 343), were eliminated because people often show preexisting response tendencies to them (Howard, Howard, Japikse, DiYani, et al., 2004; Soetens, Melis, & Notebaert, 2004). By eliminating these triplets, we could ensure that any high- versus low-probability differences were due to statistical learning and not to preexisting tendencies. We calculated mean accuracy and median RTs (for correct responses) for each participant and each epoch, separately for high- and low-probability triplets. The mean accuracy was 95.19% (*SD* = 0.03%) in the Learning Phase of the task. Since high accuracy scores and the relatively low variance in samples of neurotypical participants can hinder the detection of learning (Vékony et al., 2020), we considered RTs to be a more appropriate measure of performance in the ASRT task. Therefore, we use RTs as our primary measures in this paper. Statistical learning scores were calculated as the difference in RTs between high- and low-probability triplets (i.e., RTs for low-probability triplets minus RTs for high-probability triplets). Higher scores indicated better learning/memory performance. General skill knowledge was defined as a general decrease in median RTs during practice (i.e., participants became faster throughout the task), irrespective of triplet types. Median RTs were calculated separately for each epoch in each phase.

To evaluate statistical learning, we conducted repeated measures analyses of variance (ANOVAs) by contrasting statistical learning scores across the Learning Phase. To test general skill learning, we contrasted median RTs across the Learning Phase using repeated measures ANOVAs. As the main goal of the present paper is to investigate the consolidation of statistical and general skill knowledge, we only briefly report the results on learning in the main text and report the exact statistics in the Supplementary Materials. To evaluate the consolidation of the acquired statistical knowledge, we conducted

ANOVAs by contrasting statistical learning scores of the last epoch of the Learning Phase with those of the first epoch of Testing Phase. To evaluate the consolidation of general skill knowledge, we conducted repeated measures ANOVAs by contrasting median RTs of the last epoch of the Learning Phase with those of the first epoch of the Testing Phase. Greenhouse-Geisser epsilon (ε) correction was used when necessary. Original *df* values and corrected, two-tailed *p* values (if applicable) are reported together with partial eta-squared (η_p^2) as the measure of effect size.

As children and older adults are typically respond with slower RTs overall (e.g., Juhasz et al., 2019), we conducted additional ANOVAs on standardized RTs. To control for the effect of average RT differences across age groups on learning and consolidation of knowledge, we employed two different ways of standardization: (1) ratio scores and (2) log-transformed RT data. For calculating ratio scores, we transformed the data in the following way. We divided each participants' raw RT values of each trial type and each epoch by their own median RT in the first epoch of the task (for a similar approach, see Horvath et al., 2020; Juhasz et al., 2019; Nitsche et al., 2003; Tóth-Fáber, Janacsek, et al., 2021). This way, participants' performance was around 1 at the beginning of the task and changed as the task progressed. We then calculated standardized learning and memory scores by subtracting standardized RTs for high-probability triplets from standardized RTs for low-probability triplets. Higher standardized scores indicated better learning/memory. General skill knowledge scores were standardized in an identical way: Each participants' median RTs in Epoch 4 and Epoch 5 were divided by their median RT in the first epoch of the task. For log-transformed RT data, we applied a logN transformation on the trial-based raw RT data. Then, we computed the mean of log-transformed RTs for each trial type and each epoch, separately for each participant. Log-transformed statistical knowledge scores were calculated by subtracting log-transformed RTs for high-probability triplets from log-transformed RTs for low-probability triplets. Log-transformed general skill knowledge was calculated for each epoch using the mean of trial-based log-transformed RT data. For the sake of brevity, we only refer to the results of these ANOVAs in the main text where they are relevant in comparison to the results of raw RTs, and we report the exact statistics in the Supplementary Materials.

Moreover, to explore consolidation in more detail, we fitted curves to the data of the Learning Phase and used the fitted parameters to predict statistical learning scores and general skill performance in the Testing Phase (see Adi-Japha et al., 2014; Pan & Rickard, 2015; Rickard, 2007). A linear function was fitted to the block-wise statistical learning

scores, and a power function was fitted to the block-wise general skill learning scores. Since some participants acquire statistical regularities quickly, showing high statistical learning early in the task, then maintaining their performance throughout the task, the slope of their learning trajectory is near zero. This leads to a low R^2 value even when a linear function fits the data well. Residual standard errors (RSEs) are independent of the slope; therefore, they are better goodness-of-fit estimates in these cases. Hence, we report RSEs instead of R^2 values, both for the statistical learning and general skill learning scores, for comparability. Smaller RSEs indicate better fit in both cases.

Importantly, in previous studies that used curve fitting on learning data, a performance improvement (i.e., offline learning) was typically expected after an offline delay (Adi-Japha et al., 2014; Pan & Rickard, 2015; Rickard, 2007). In these cases, using fitted parameters from a power or a linear function has been an appropriate approach to predict and test future performance (Pan & Rickard, 2015). Therefore, using curve fitting to test offline changes can work well for general skill learning in our study because offline learning is expected following the 24-hour delay. However, this approach may be less ideal for statistical learning as measured by the ASRT task because maintenance of performance (i.e., retention) may be expected instead of offline learning (e.g., Kóbor et al., 2017; Simor et al., 2019; Song et al., 2007b). Moreover, differences in variance across age groups can create additional challenges when curve fitting is used to test age-related differences in learning and consolidation: As variance can influence how well a function fits the data and how reliable the predicted performance is, differences in variance can hinder the comparability of predicted performance across age groups. Nevertheless, we report the curve fitting results to provide a more detailed picture of the consolidation of statistical and general skill knowledge across the lifespan using multiple approaches.

In conjunction with the frequentist analyses, we performed Bayesian mixed-design ANOVAs and Bayesian paired-samples t-tests for the relevant comparisons. Bayesian mixed-design ANOVAs were run on the memory scores to test which factors determine performance. Here, we present Bayesian Model Averaging and the exclusion Bayes Factor ($BF_{exclusion}$). $BF_{exclusion}$ values quantifies the change from prior to posterior odds and can be interpreted as the evidence in the data for excluding a given predictor from the model. Thus, values below 1 support the inclusion and values above 1 the exclusion of the given factor. Cauchy prior distribution was used for the ANOVA with a fixed-effects scale factor of $r = 0.5$, and a random-effects scale factor of $r = 1$. Moreover, we ran Bayesian paired-samples t-tests for comparing performance between the end of

the Learning Phase and the beginning of the Testing Phase, separately for each age group. Bayes factor (BF) was computed to assess the amount of evidence for the null-hypothesis of no offline change. The BF is a statistical technique that helps conclude whether the collected data favors the null-hypothesis (i.e., evidence for no difference between groups or variables) or the alternative hypothesis (i.e., evidence for differences); thus, the BF could be considered as a weight of evidence provided by the data (Wagenmakers et al., 2011). One of the main benefits of calculating the BF is that for non-significant comparisons we can use the BF to conclude that the acquired evidence supports H_0 rather than H_1 (Dienes, 2011, 2014; Wagenmakers, 2007). BFs were calculated using JASP version 0.14 (Rouder, Speckman, Sun, Morey, & Iverson, 2009). Here we report BF_{01} values where greater values support the null-hypothesis (no difference) over the alternative hypothesis. According to Wagenmakers et al. (2011), BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis.

3. Results

3.1. Are there age-related differences in the consolidation of statistical knowledge?

Before testing the age-related differences in consolidation of statistical knowledge, we tested the potential age-related differences in statistical learning. Analysis on raw RT data in the Learning Phase showed better learning under the age of 13, whereas analysis on ratio scores revealed comparable learning from childhood to young adulthood, followed by decreased learning from the age of 30. We present the exact statistics in the Supplementary Materials.

To test 24-hour consolidation of the acquired statistical knowledge, we contrasted statistical learning scores computed for the last epoch of the Learning Phase (Epoch 4) with the learning scores computed for the first of the Testing Phase (Epoch 5) and submitted these scores to a mixed-design ANOVA with EPOCH (Epoch 4 vs Epoch 5) as a within-subject factor and AGE GROUP as a between-subjects factor. The ANOVA revealed overall significant statistical knowledge (main effect of INTERCEPT: $F(1, 246) = 309.24$, $p < 0.001$, $\eta_p^2 = 0.56$) and significant differences in overall learning across age

groups (main effect of AGE GROUP: $F(8, 246) = 2.91$, $p = 0.004$, $\eta_p^2 = 0.09$). Importantly, statistical knowledge appears to be retained over the 24-hour delay period with no significant change between the end of the Learning Phase and the Testing Phase (main effect of EPOCH: $F(1, 246) = 0.39$, $p = 0.53$, $\eta_p^2 = 0.002$). Moreover, no age group differences emerged in the retention of the statistical knowledge (non-significant EPOCH × AGE GROUP interaction: $F(8, 246) = 0.14$, $p = 0.997$, $\eta_p^2 = 0.005$; all ps > .52): this suggests that all age groups retained the acquired knowledge over the 24-hour delay period (Fig. 5.2 and Fig. S5.1). The analysis of effects of the Bayesian mixed-design ANOVA showed that the main effect of EPOCH and the EPOCH × AGE GROUP interaction should be excluded from the model (see Table 5.2, and for model comparisons, see Table S5.1), corroborating the findings of the frequentist ANOVA.

To rule out the possible confounding effect of different average RTs across age groups on these results, we employed two ways of standardization, and we conducted two ANOVAs for ratio scores and log-transformed RT data, respectively (for details on the standardization process, see the Statistical analysis section). We submitted the standardized statistical learning scores to mixed-design ANOVAs with EPOCH (Epoch 4 vs Epoch 5) as a within-subject factor and AGE GROUP as a between-subjects factor. The ANOVAs revealed identical results to the ANOVA computed on raw RT scores (for the exact statistics, see Table S5.7 and the accompanying text). The Bayesian mixed-design ANOVAs on the standardized learning scores also supported these findings (for details, see Table S5.3-S5.4 and Table S5.8-S5.9), confirming no change in learning scores over the 24-hour delay period and no differences in this pattern across age groups.

To explore consolidation in more detail, we used a linear function to predict performance in the Testing Phase (for details, see the Statistical analysis section). RSEs were calculated separately for the age groups, and they were between 3.80 and 19.17 (M = 8.41), suggesting a generally good fit to the data. A difference score was calculated by subtracting the predicted statistical learning scores from the observed statistical learning scores. We submitted the difference scores to a mixed-design ANOVA with BLOCK (1-5) as a within-subject factor and AGE GROUP as a between-subjects factor. Importantly, we found no significant age-related effect in this ANOVA: the difference score between the predicted and observed statistical learning scores was comparable across the age groups. Bayesian ANOVA further supported these results. For the exact statistics, see Table S5.13-S5.15.

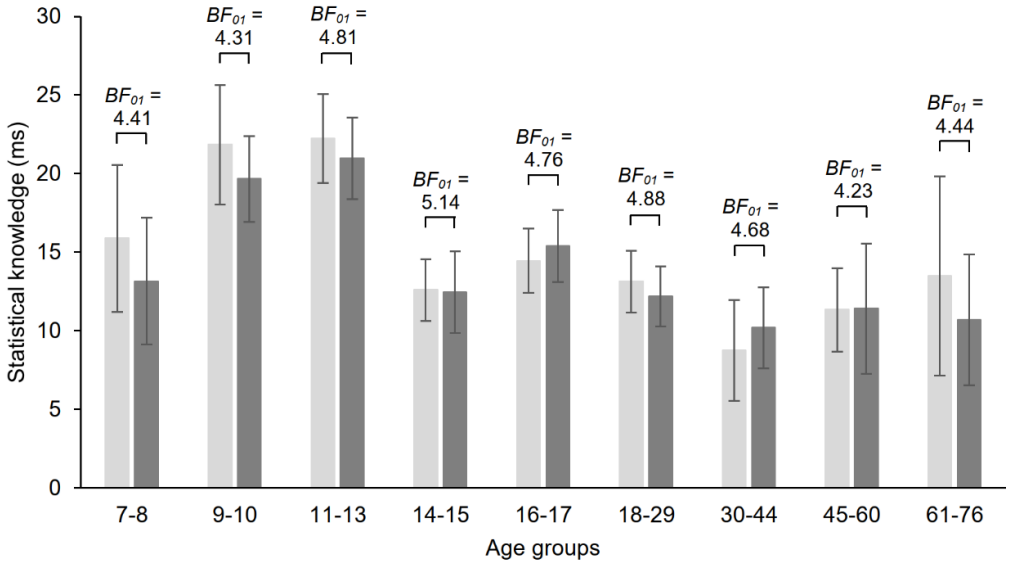

Figure 5.2. Consolidation of statistical knowledge over the 24-hour offline period across age groups. RT **s**tatistical learning scores for the last epoch of the Learning Phase (Epoch 4, light gray bars) were contrasted with those for the first epoch of the Testing Phase (Epoch 5, dark gray bars). BF_{01} values were obtained by paired-samples t-tests for this contrast separately for each age group. All reported BF_{01} values indicate substantial evidence for the null-hypothesis ($BF_{01} > 3$), providing evidence for comparable knowledge in Epoch 4 and Epoch 5 in each age group. Error bars denote the standard error of mean (SEM).

Table 5.2. Analysis of effects of Bayesian ANOVA for consolidation of statistical knowledge.

Effects	**P(incl)**	**P(incl|data)**	**$BF_{exclusion}$**
Epoch	0.400	0.104	8.608
Age group	0.400	0.800	0.250
Epoch × Age group	0.200	3.382e-4	246.000

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

3.2. Are there age-related differences in the consolidation of general skill knowledge?

Similar to statistical learning, before comparing the age groups on the consolidation of general skill knowledge, we first tested the age-related differences in general skill learning. Analysis on raw RT data in the Learning Phase revealed highest general skill learning performance in the youngest age group. For details, see the exact statistics in the Supplementary Materials.

We tested the consolidation of general skill knowledge (defined as median RTs changes) over the delay period with a mixed-design ANOVA on median RTs (i.e., RTs irrespective of the probabilities of events) with EPOCH (Epoch 4 vs Epoch 5) as a within-subject factor and AGE GROUP as a between-subjects factor. Our analysis found that the median RTs significantly decreased over the 24-hour delay (main effect of EPOCH: $F(1, 246) = 107.92$, $p < 0.001$, $\eta_p^2 = 0.31$): participants responded faster in the Testing Phase compared to the end of the Learning Phase (significant speed-up in all age groups: all *p*s < 0.014, except for the 14-15-year-old group, where $p = 0.080$, Fig. 5.3). The amount of speed-up over the delay period, however, was not uniform across the age groups (EPOCH × AGE GROUP interaction: $F(8, 246) = 2.26$, $p = 0.02$, $\eta_p^2 = 0.07$). A follow-up ANOVA on the offline change score (i.e., RTs in Epoch 4 minus RTs in Epoch 5) showed that the 7-8-year-olds exhibited the greatest speed-up over the delay, significantly differing from the speed-up of almost all other age groups (*p*s < 0.026; 7-8-year-old vs 9-10-year-old groups: $p = 0.068$; Fig. S5.2). The other age groups' median RT changes over the delay period were not significantly different from one another (all *p*s > 0.062). Bayesian mixed-design ANOVA also supported the inclusion of the main effect of EPOCH and the EPOCH × AGE GROUP interaction (Table 5.3, and for model comparisons, see Table S5.2).

Similarly to the statistical learning scores, we ran two ANOVAs on standardized RTs as well, one for ratio scores and one for log-transformed RT data (for details on the standardization process, see the Statistical analysis section). Both ANOVAs revealed a significant RT speed-up over the delay period, however, in contrast to the ANOVA on raw RT scores, the age groups did not differ from each other in the amount of speed-up either concerning ratio scores or log-transformed RT data (for the exact statistics, see Table S5.10, Figure S5.3 and the accompanying text). The Bayesian mixed-design ANOVAs also supported the inclusion of the main effect of EPOCH and the exclusion of the EPOCH × AGE GROUP interaction, suggesting a uniform speed-up over the delay period across the age groups (for details, see S5.5-S5.6 and S5.11-S5.12). These results suggest that the group differences observed in the raw average RT analyses above were largely driven by some age groups being on average slower in the task than other groups; controlling for this confound eliminated the group differences in the consolidation of general skill knowledge over the delay period.

Similarly to the statistical learning scores, we further examined the magnitude of offline gains and possible age-related differences by predicting future general skill

performance in the Testing Phase with a power law function (for details, see the Statistical analysis section). RSEs were calculated separately for the age groups, and they were between 4.32 and 20.54 (M = 10.53), suggesting a generally good fit to the data. A difference score was calculated by subtracting the predicted RT data from the observed RT data. We submitted the difference scores to a mixed-design ANOVA with BLOCK (1-5) as a within-subject factor and AGE GROUP as a between-subjects factor. We found no significant age-related effect: the difference between the predicted and observed performance was comparable across the age groups. Bayesian ANOVA further supported these results. For the exact statistics, see Table S5.16-S5.18.

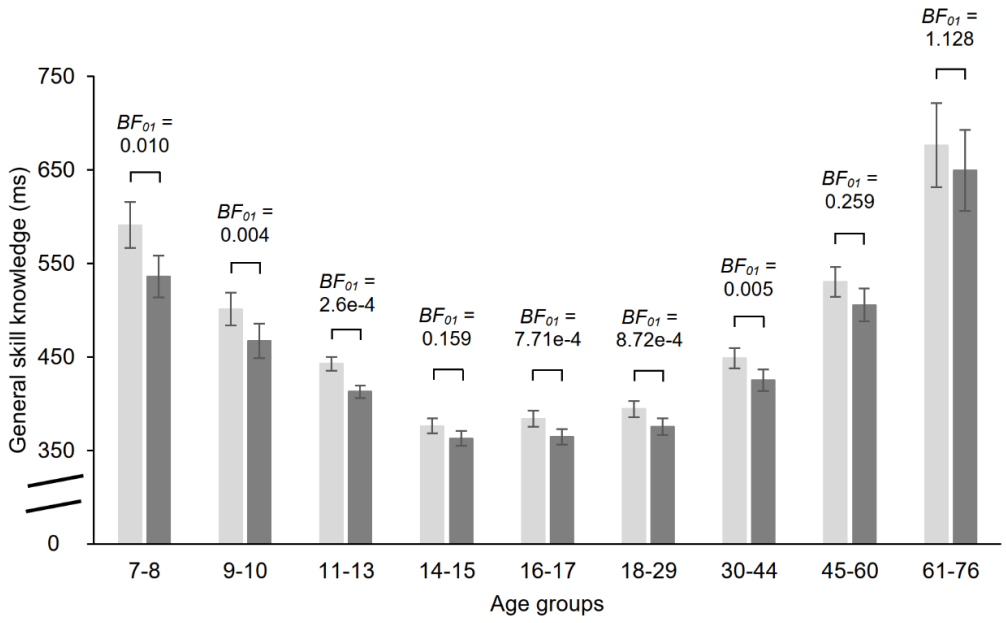

Figure 5.3. Consolidation of general skill knowledge over the 24-hour offline period across age groups. Average RT values for the last epoch of the Learning Phase (Epoch 4, light gray bars) were contrasted with those for the first epoch of the Testing Phase (Epoch 5, dark gray bars). BF_{01} values were obtained by paired-samples t-tests for this contrast separately for each age group. BF_{01} values for all age groups except for the 61-76-year-olds indicate substantial evidence for the alternative hypothesis (BF_{01} < 0.33) providing evidence for offline learning over the 24-hour delay. BF_{01} value obtained for the 61-76-year-olds could not provide evidence for either the null or the alternative hypotheses. Error bars denote the SEM.

Table 5.3. Analysis of effects of Bayesian ANOVA for consolidation of general skill knowledge.

Effects	P(incl)	P(incl|data)	$BF_{exclusion}$
Epoch	0.400	0.548	1.067e -17
Age group	0.400	0.548	5.163e -29
Epoch × Age group	0.200	0.452	1.214

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

3.3. Testing possible confounds influencing the consolidation of statistical and general skill knowledge

To test the possible confounding effect of averaging over the last five blocks of the Learning Phase and the first five blocks of the Testing Phase (Pan & Rickard, 2015), we contrasted performance in the last block of the Learning Phase (Block 20) and in the first block of the Testing Phase (Block 21). The analysis revealed that some degree of forgetting might be present as statistical learning score in Block 21 was marginally lower than that in Block 20. Bayesian analysis was not conclusive, the Bayes Factor was around 1, not supporting either the null or the alternative hypothesis. Crucially, age-related differences were not detected either with the frequentist or with the Bayesian analysis. For the exact statistics, see Table S5.19-S5.20 and the accompanying text.

It is important to note that the analysis of block-wise data in the ASRT task should be interpreted carefully due to the relatively low number of trials. Specifically, statistical learning scores are calculated as difference scores between the high- and low-probability trials after excluding the first five random practice trials at the beginning of the block, erroneous responses as well as trills and repetitions from the 85 trials that are presented in a block. Hence, aggregated (mostly epoch-level) data has been used to characterize learning in the ASRT task since its inception because it enables to track the trajectory of learning while simultaneously decreasing the effect of noise in the learning scores to an acceptable level (Howard, & Howard, 1997; Song et al., 2007a, 2007b).

Moreover, to test whether practice-dependent changes influenced the offline learning of general skill knowledge over the 24-hour offline delay, we compared performance in the last block of the learning phase (Block 20) and the first block of the Testing Phase (Block 21). Overall, participants responded faster in Block 21 compared to Block 20, with similar speed-up across the age groups, and the amount of speed-up was

comparable across the age groups. This suggests that the offline learning of general skill knowledge over the delay period was not due to further practice-dependent changes in the Testing Phase. For the exact statistics of this analysis, see Table S5.21 and the accompanying text.

We also tested whether offline learning of general skill knowledge in terms of RTs could be influenced by decreased accuracy over the offline period. On the group level, mean accuracy scores increased over the offline delay, but the level of improvement was not comparable across the age groups. Significant offline learning was only detectable in the 7-8-year-old and 11-13-year-old groups. These results suggest that, in terms of accuracy scores, none of the age groups showed forgetting, therefore, offline learning in terms of RTs cannot be explained by a decreased accuracy over the offline period. For exact statistics, see Supplementary Materials.

4. Discussion

The present study examined the 24-hour consolidation of statistical and general skill knowledge in a large sample of participants between the age of 7 and 76 using the same experimental design across the sample. Based on statistical learning scores computed from raw RT data, we showed retained statistical knowledge in all age groups. Analyses on standardized RT data and Bayesian analyses (both on raw RTs and standardized RTs) further corroborated these results. As for general skill knowledge, while the analyses on raw RT data suggested offline improvement that was the greatest in the 7-8-year-olds, results on standardized RT data revealed offline gains in all age groups with a uniform speed-up across the sample. Bayesian analyses of general skill consolidation also confirmed this uniform speed-up.

Our results on age-invariance in the retention of statistical knowledge are in line with prior smaller scale studies that focused on one age group only or contrasted performance in a few age groups. Specifically, in groups of children and adolescents, successful retention of statistical regularities has been shown following a 16-hour (Nemeth, Janacsek, Balogh, et al., 2010; Takács et al., 2018), 24-hour (Hedenius et al., 2013), three-day (Hedenius et al., 2011) and even a one-year offline period (Tóth-Fáber, Janacsek, et al., 2021). Offline learning has also been found over a 24-hour delay in a group of 9-13-year-old participants (Hedenius et al., 2020). Successfully retained knowledge had been consistently demonstrated in young and middle adulthood as well

over various offline delays ranging from hours to even one year (e.g., Arciuli & Simpson, 2012; Kóbor et al., 2017; Meier & Cock, 2014; Nemeth, Janacsek, Király, et al., 2013; Romano et al., 2010).

However, differences in consolidation between children and adults have been suggested in specific areas, such as the contribution of sleep to consolidation. Fischer et al. (2007) compared the consolidation of statistical knowledge in 7-11-year-old children and young adults after offline periods of overnight sleep or daytime wakefulness. Adults showed better retention of statistical knowledge following sleep compared to the wakeful offline period, whereas children showed an opposite pattern with better retention after daytime wakefulness than after overnight sleep. Notably, children showed overall higher learning than adults and this difference in the pre-sleep performance level could have strongly confounded the results on consolidation (Wilhelm et al., 2012). Our study was not designed to test the effect of sleep on consolidation across the lifespan: the 24-hour delay employed in our study included both periods of overnight sleep and daytime wakefulness. In theory, an opposite pattern such as the one observed in the study of Fischer et al. (2007) could have resulted in overall similar retention performance of statistical knowledge across age groups. Importantly, we chose the 24-hour delay because that could provide a clearer picture of everyday functioning of individuals as it incorporates both the effects of sleep and daytime wakefulness. Overall, consolidation of statistical knowledge seems to be similar across childhood and adulthood, and previously found developmental differences might be explained by confounding factors, such as differences in sleep.

Previous studies investigating the consolidation of statistical knowledge in aging showed somewhat mixed results. Most studies suggested intact consolidation of statistical knowledge in elderly adults. Comparable retention has been shown in younger and older adults after a 12-hour offline period, irrespective of whether that period included overnight sleep or daytime wakefulness (Nemeth, Janacsek, Londe, et al., 2010). Retained knowledge has also been found following a one-year delay (Romano et al., 2010). In contrast, Nemeth and Janacsek (2011) have found forgetting of statistical knowledge in older adults compared to successful retention in young adults, irrespective of the length of the offline delay (12-hour, 24-hour or one-week). Our results align with the former studies, providing substantial evidence for retained statistical knowledge over the 24-hour delay in older adults, as indicated by the Bayesian analyses (BF_{01} = 4.44). The mixed findings across studies could be attributed to differences in experimental designs, for

example, different lengths of the learning period could lead to a varying degree of fatigue, which can potentially affect consolidation. Future studies may systematically test how such differences affect consolidation across age groups.

During learning, we extract the regularities from the environment and encode them into initially fragile memory representations. During consolidation, the recently acquired representations undergo a progressive stabilization, creating long-term memory representations (Walker, 2005). Although these two processes are intertwined and depend on each other, considering the age-variant lifespan trajectory of statistical learning based on previous studies (Janacsek et al., 2012; Juhasz et al., 2019; Lukács & Kemény, 2015) and the age-invariant lifespan trajectory of consolidation of statistical knowledge based on the present study, we can conclude that these two processes show a dissociation. Similar results have been demonstrated before in atypical development. In Tourette syndrome, intact consolidation has been shown to accompany enhanced learning (Takács et al., 2018), whereas intact learning and impaired consolidation have been found in developmental dyslexia (Hedenius et al., 2021). Our results are in line with these prior ones as we showed a dissociation between learning and consolidation of statistical knowledge in a neurotypical population: while learning varied with the function of age, consolidation did not. This suggests that, at least partially, distinct mechanisms and neural networks underlie the acquisition and consolidation of statistical knowledge across the lifespan.

To the best of our knowledge, no theoretical framework has been proposed for the lifespan trajectory of consolidation of statistical knowledge. Here, we found comparable retention of statistical knowledge in all ages from 7 to 76 years, which would suggest an age-invariance model of consolidation of such knowledge. Based on the present behavioral results, we can also make assumptions for the neural networks that underlie the consolidation of statistical knowledge, at least when measured with the ASRT task. As mentioned in the introduction, in their model for statistical learning, Janacsek et al. (2012) proposed a shift from detecting raw probabilities to relying more on internal interpretations of events, which then leads to decreased statistical learning performance in adults compared to children under the age of 12. On the neurobiological level, the competition model (Janacsek et al., 2012) related this shift to the protracted maturation of the hippocampus and prefrontal cortex. While these brain regions have been suggested to underlie the development of internal models, basal ganglia, particularly the striatum, have been linked to the detection of raw probabilities. Albouy, King, Maquet, and Doyon

(2013) proposed a model for a related process, that is, motor sequence learning and consolidation. According to this model, in healthy young adults, during the acquisition of a motor sequence, the hippocampus and the striatum show an antagonistic dynamic, which is presumably mediated by the prefrontal cortex. In contrast, during consolidation and retest, instead of a competitive dynamic, the striatum and the hippocampus seem to function cooperatively. According to Albouy et al. (2013), striatal activity supports time-dependent maintenance in performance (i.e., retention), whereas the hippocampus supports sleep-dependent enhancement in performance (i.e., offline learning), at least in young adults. In line with this dissociation, Schapiro et al. (2019) also showed the involvement of the hippocampus in the sleep-dependent consolidation of a motor sequence in middle aged adults. Converging our behavioral results and the model of Albouy et al. (2013), we can speculate that consolidation of statistical knowledge was likely more reliant on the striatum and its circuits as we found retention of statistical knowledge in all age groups. This would be consistent with the notion that the consolidation of statistical knowledge, at least when measured with the ASRT task, is independent of sleep (Nemeth, Janacsek, Londe, et al., 2010; e.g., Song et al., 2007b). It is important to note that in the present study, we employed a visuomotor task involving temporally distributed non-adjacent statistical regularities, therefore the conclusion of age-invariance is restricted to these regularities. Considering other, related processes, such as motor (sequence) learning, usually measured with deterministic SRT and finger tapping tasks, a different trajectory might emerge. As stated above, the consolidation of motor (sequence) knowledge appears to be sleep-dependent, resulting in offline gains over the delay, due to the involvement of the hippocampus (Albouy et al., 2008). Due to sleep-dependency and the underlying neural circuits (i.e., the hippocampus showing protracted development (Keresztes, Ngo, Lindenberger, Werkle-Bergner, & Newcombe, 2018)), age-variance may be expected for the consolidation of motor (sequence) knowledge. Indeed, empirical evidence supports this notion (Prehn-Kristensen et al., 2009; Roig, Ritterband-Rosenbaum, Lundbye-Jensen, & Nielsen, 2014; Urbain, Houyoux, Albouy, & Peigneux, 2014, but see Pan & Rickard, 2015; Rickard & Pan, 2017). In contrast, the time-dependent consolidation of statistical knowledge seems to be age-invariant, potentially due to the greater reliance on the striatum that matures early in development. To sum up, the cooperation of the striatum and the hippocampus could be responsible for the consolidation of the acquired knowledge and based on our results, the striatum may be more prominent in this interaction when consolidation is independent of

sleep, as is the case for the statistical knowledge tested in the present study. Nevertheless, it is important to note that this is highly tentative and further neuroimaging studies are needed to corroborate it.

Previous development and aging studies have highlighted the importance of baseline RT differences across the lifespan. It is well-established that children and older adults show slower RTs compared to young adults (e.g., Janacsek et al., 2012; Juhasz et al., 2019). These differences can confound performance as children and older adults have more room to improve, meaning that as their baseline RTs are slower, they can demonstrate higher gains in learning. Juhasz et al. (2019) focused on general skill learning in the sample used in the present study. According to their results, general skill learning is heightened in childhood, but this could not be explained by the 'more room to improve' concept. Superior general skill learning in 7-8-year-olds persisted across different analysis approaches which controlled for the baseline speed differences. Our results on raw RT data showed that 7-8-year-olds are also superior in the consolidation of general skill knowledge as they exhibited greater offline improvement than the other age groups. Importantly, however, this could not be confirmed by the analyses of standardized RT data where we controlled for the average speed differences across the age groups. Hence, the greater offline speed-up compared to other age groups seen on raw RT data is possibly due to the generally slower responses in the 7-8-year-olds. Thus, while general skill learning seems to be age-variant with heightened learning in childhood (Juhasz et al., 2019), consolidation of such knowledge seems to be age-invariant with similar offline improvement in all age groups, based on the results of the current paper.

Prior studies have found inconsistent results on the consolidation of general skill knowledge in the elderly population. Based on these studies, the length of the offline delay seems to influence the magnitude of consolidation in older adults (Nemeth & Janacsek, 2011; Nemeth, Janacsek, Londe, et al., 2010; Romano et al., 2010). Concerning the prior studies, only Nemeth and Janacsek (2011) employed a 24-hour offline delay. They have found neither offline improvement nor forgetting in older adults, general skill knowledge did not change over the offline period. In contrast, we found offline improvement in 61-76-year-olds following a 24-hour delay and the magnitude of offline gains did not differ significantly between younger and older adults. However, it is worth noting that Bayesian evidence for offline gains in general skill knowledge was around 1 in the 61-76-year-old group, which means that we did not find evidence for either the null

or the alternative hypothesis. Hence, further research is warranted to explore the consolidation of general skill knowledge in older adulthood.

A considerable amount of research has focused on the changes of different cognitive functions across the lifespan. The lifespan trajectory of several cognitive functions has been described as an inverted U-shape: these functions continuously mature through childhood and adolescence, then peak in young or middle adulthood and decline over the course of aging (e.g., Alloway & Alloway, 2013; Zelazo et al., 2004). Some functions peak in childhood and starts to decline as soon as adolescence or young adulthood (e.g., Janacsek et al., 2012; Johnson & Newport, 1989; Juhasz et al., 2019), whereas some functions remain intact in late adulthood as well (e.g., Ikier et al., 2008). The consolidation of statistical regularities seems to follow a different, age-invariant trajectory: the acquired statistical knowledge is comparably retained in all age groups from the age of 7 to 76 years. Consolidation of general skill knowledge also seems to follow an age-invariant trajectory: in this case, the offline improvement over the delay period is comparable across all age groups. Since the oldest adult in our study was 76 years old, future studies could provide further insights into how aging affects consolidation by involving individuals beyond this age as well.

One limitation of our study is the lack of interference design, which can be important for consolidation studies (Brashers-Krug, Shadmehr, & Bizzi, 1996; Ellenbogen, Hulbert, Stickgold, Dinges, & Thompson-Schill, 2006). Consolidation can be defined both as stabilization of the acquired knowledge—usually evident as no change in performance or offline learning during the delay period—and as resistance to interference. A previous study with the ASRT task employed an interference design: Kóbor et al. (2017) investigated the one-year consolidation of statistical knowledge in healthy young adults. They showed successful retention as well as a resistance to interference. Hence, successful consolidation was expressed both by retention and resistance to interference in the same group of participants. Employing an interference design might also be relevant from a developmental perspective. Dorfberger, Adi-Japha, and Karni (2007) showed similar consolidation in childhood and adulthood on the behavioral level using a motor learning task, however, a difference emerged concerning interference. Children seemed to be less susceptible to subsequent interference compared to adults, suggesting age-related differences in this aspect of consolidation. Therefore, future studies should employ an interference design to shed further light on the lifespan trajectory of memory consolidation.

Conducting large-scale studies involving participants from age groups across the lifespan is essential to characterize the development and aging of any cognitive functions. Inferences made from different sets of individual studies have a risk of different paradigms and designs confounding the conclusions. Here, we employed a lifespan approach in a cross-sectional design involving participants from 7 to 76 years using the same task across the whole sample. In conclusion, the present study demonstrated comparable consolidation of statistical and general skill knowledge following a 24-hour offline delay across the lifespan, from 7-year-olds to 76-year-olds. Our study offers evidence for age-invariance in these key cognitive functions.

VI. General discussion

The dissertation aimed to investigate how the changes in fronto-striatal networks – considering both typical and atypical development – lead to changes in procedural memory. Across the presented four studies, we investigated this question from different viewpoints. In Study 1, we assessed a childhood-onset neurodevelopmental disorder with alterations in the fronto-striatal networks, namely Tourette syndrome, and explored different regularities within the procedural memory system. While Study 1 focused on the acquisition of such regularities, in Study 2, we went one step further and examined whether retention changes as a result of altered fronto-striatal networks. Thus, Study 2 examined the short- and long-term consolidation of different regularities in Tourette syndrome. To get a more detailed picture of the procedural memory consolidation processes, Study 3 tested the one-year retention of different regularities in typically developing children and adolescents. Finally, to gain a deeper understanding of developmental processes, in the last study, we went beyond past research involving only a few age groups (i.e., children vs. young adults or young vs. old adults) by taking a lifespan approach. Hence, Study 4 aimed to shed light on the lifespan differences in procedural knowledge consolidation in a large sample of participants aged 7 to 76 years. In the next section, we summarize the main findings of each study.

1. Summary regarding atypical development

1.1. Is the learning of probability-based and serial order-based regularities enhanced in Tourette syndrome?

In Study 1, we examined how probability-based and serial order-based regularities within the procedural memory system are affected in TS and whether they contribute to the possible procedural hyperfunctioning proposed by previous studies (Dye et al., 2016; Shephard et al., 2019; Takács et al., 2018; Walenski et al., 2007). Employing the cued version of the ASRT task made it possible to investigate these two regularities simultaneously. Considering probability-based regularities, we found enhanced sensitivity in TS: children with TS extracted these regularities faster than their typically developing peers. In contrast, the acquisition of serial order-based regularities was impaired in TS: children with TS did not learn these regularities, whereas typically developing participants did.

Enhanced sensitivity to probability-based regularities in TS is in line with the prior studies proposing procedural hyperfunctioning in the disorder (Shephard et al., 2019; Takács et al., 2018). Takács et al. (2018) found enhanced learning of statistical regularities in TS employing the uncued, original version of the ASRT task. Hence, a speeded extraction of such regularities in TS is present in the uncued and cued version of the task as well. Shephard et al. (2019) also suggested procedural hyperfunctioning in TS: children with TS showed difficulties in the transition from sequence to non-sequence blocks in a serial reaction time task.

The enhanced sensitivity to probability-based regularities in TS also falls in line with prior findings suggesting procedural hyperfunctioning in the disorder using language-based tasks (Dye et al., 2016; Walenski et al., 2007). The extraction of complex statistical regularities is fundamental in language acquisition and processing, such as detecting word boundaries (Saffran et al., 1996) and phrase boundaries (Thompson & Newport, 2007). In the study of Dye et al. (2016), participants solved a non-word repetition task, which involves the rule-governed decomposition and recomposition of the words, which is influenced by the phonotactical constraints of the language and, therefore, by statistical regularities (Coady & Evans, 2008). Walenski et al. (2007) showed empirical evidence for the speeded processing of regular, rule-governed past tenses in TS. In contrast, the production of irregular past tenses was comparable between the TS and control groups. The production of rule-governed past tenses is also related to statistical regularities. Enhanced procedural functions have manifested as speeded tool naming as well, while reaction times for naming animals were similar in the TS and control groups. Naming objects is not clearly related to statistical regularities, but it is possible that enhanced statistical learning has an additive effect, which results in speeded tool naming.

The impairment of the acquisition of serial order-based regularities in TS is also not without precedent. Avanzino et al. (2011) employed a finger-tapping task with a deterministic sequence and showed impaired performance in TS children. Palminteri et al. (2011) employed high and minimal reward conditions in a motor learning task and showed enhanced learning in the high reward condition and impaired learning in the minimal reward condition. As our task did not involve rewards, our result showing impaired learning of serial order-based regularities is in line with the results of Palminteri et al. (2011) on impaired learning in the minimal reward condition.

1.2. Is the consolidation of probability-based and serial order-based regularities robust in Tourette syndrome?

The goal of Study 2 was to examine the consolidation of probability-based and serial order-based regularities over a short-term (five-hour) and long-term (one-year) offline delay in children with TS and typically developing peers. Employing a one-year-long offline period is rare in laboratory studies, and incorporating a long offline period enhances the ecological validity of the findings: acquiring and retaining skills happen over a longer time than the usual time scale of a lab visit. Here, both groups showed successful retention of probability-based regularities following the offline periods and retention performance was comparable between the groups. The retention of serial order-based regularities could not be investigated in this study as children with TS did not acquire these regularities; therefore, their consolidation could not be reliably tested here.

Successful one-year retention of probability-based regularities has been shown in neurotypical adults (Kóbor et al., 2017; Romano et al., 2010) and in typically developing children as well (Study 3 of the present dissertation). In TS, prior to our study, only the short-term, 16-hour retention was investigated: Takács et al. (2018) used the uncued version of the ASRT task and in conjunction with enhanced statistical learning, greater forgetting has been shown. Importantly, as the authors also suggest, no firm conclusions can be drawn from this study due to the learning differences between the TS and control groups. After controlling for learning differences, the groups showed comparable retention of statistical knowledge. To sum up, we replicated the results of Takács et al. (2018) on (relatively) short-term retention and went beyond this study by showing that long-term (one-year) retention is also robust in TS.

We need to note the discrepancy in the results of Study 1 and 2. In Study 2, we found intact, but not enhanced learning of probability-based regularities in TS, which is somewhat different compared to the findings of Study 1, where we found a speeded extraction of such regularities. Importantly, in Study 2, we used a subset of the TS children from Study 1, but the control groups differed due to difficulties assessing the original control group one year later. Throughout the studies of the dissertation, we used Bayesian analyses to check whether our null findings (i.e., no group differences) were corroborated by Bayes Factors or due to the lack of power. We did not run Bayesian analyses when conducting Study 1, but we have now filled this gap here by analyzing the data this way as well. Bayesian mixed-design ANOVAs were run separately for the learning scores of probability-based regularities and serial-order regularities. Regarding

probability-based regularities, anecdotal evidence was found for the Epoch × Group interaction (Table 6.1 and 6.2), supporting the results of the frequentist analyses (pp. 29). As for serial-order regularities, the lack of learning in the TS group was corroborated by the Bayesian analyses as shown by the anecdotal evidence for the main effect of Group (Table 6.3 and 6.4). Nevertheless, the lack of clear replication of the enhanced sensitivity to probability-based regularities suggests that we need to handle these results with caution and future studies are warranted to replicate and further examine these questions.

Table 6.1. Analysis of effects of Bayesian ANOVA for learning of probability-based regularities.

Effects	P(incl)	P(incl|data)	$BF_{inclusion}$
Epoch	0.400	0.079	0.092
Group	0.400	0.337	0.562
Epoch × Group	0.200	0.065	2.261

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{inclusion}$ column shows the inclusion Bayes Factors. $BF_{inclusion}$ values below 1 support the exclusion and values above 1 the inclusion of the given factor.

Table 6.2. Bayesian model comparisons for learning of probability-based regularities.

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.549	4.864	1.000	
Group	0.200	0.308	1.780	0.561	2.015
Epoch + Group + Epoch × Group	0.200	0.065	0.276	0.118	2.606
Epoch	0.200	0.050	0.211	0.091	0.489
Epoch + Group	0.200	0.029	0.118	0.052	1.413

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{10} column the Bayes factors for each model compared to the null model. BF_{10} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_1. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_0. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

Table 6.3. Analysis of effects of Bayesian ANOVA for learning of serial-order regularities.

Effects	P(incl)	P(incl|data)	$BF_{inclusion}$
Epoch	0.400	0.152	0.185
Group	0.400	0.664	2.146
Epoch × Group	0.200	0.026	0.254

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{inclusion}$ column shows the inclusion Bayes Factors. $BF_{inclusion}$ values below 1 support the exclusion and values above 1 the inclusion of the given factor.

Table 6.4. Bayesian model comparisons for learning of serial-order regularities.

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.261	1.409	1.000	
Group	0.200	0.561	5.109	2.153	1.034
Epoch + Group	0.200	0.103	0.461	0.397	1.701
Epoch	0.200	0.049	0.206	0.188	3.646
Epoch + Group + Epoch × Group	0.200	0.026	0.108	0.101	1.534

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{10} column the Bayes factors for each model compared to the null model. BF_{10} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_1. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_0. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

2. Summary regarding typical development

To discover any potential developmental impact on procedural memory consolidation processes, Study 3 and 4 focused on typical development, examining 9-15-year-old children and adolescents in Study 3 and employing a lifespan approach with participants aged between 7 and 76 years in Study 4. These findings could help develop a theoretical model for age-related changes in procedural memory consolidation.

2.1. Is the one-year consolidation of probability-based and serial order-based regularities successful in typically developing children and adolescents?

Study 3 examined the one-year consolidation of probability-based as well as serial order-based regularities in a sample of children and adolescents aged 9-15 years. We chose a

one-year offline period to overcome the limitation rooted in the usual time scales of laboratory studies and to enhance the ecological validity of our findings. Retained knowledge of both regularities has been shown: participants successfully acquired the regularities, which then were resistant to forgetting over a one-year delay. These results were corroborated by Bayesian statistics as well, which further strengthens the evidence for successful one-year retention. Our results are in line with previous studies using shorter delays varying from hours to 1-week. For probability-based knowledge, retention was shown following an 11-hour delay (Fischer et al., 2007). For serial order-based knowledge, Desmottes et al. (2017) demonstrated offline learning after 24-hour and 1-week delays. Using the uncued ASRT task, several studies have found retained knowledge over relatively short-term offline periods (Hedenius et al., 2011; Nemeth, Janacsek, Balogh, et al., 2010; Takács et al., 2018) and one showed offline learning after a 24-hour delay (Hedenius et al., 2021). Hence, converging evidence suggest that in typical development, retention of procedural knowledge is successful from hours to up to a year.

Combining the results of Study 3 and of prior studies (Kóbor et al., 2017; Romano et al., 2010), we can get insight into the developmental trajectory of procedural memory consolidation. We tested consolidation in a sample of 9-15-year-olds and prior studies tested consolidation using the uncued ASRT task in healthy young adults (Kóbor et al., 2017; Romano et al., 2010). Our results are in line with the findings of these studies, suggesting that retention is comparable in childhood and adulthood in typical development. Moreover, the lack of correlation between age and retention in our sample also points towards the direction of developmental invariance in procedural memory consolidation. Nonetheless, the results of Study 3 and prior studies can only provide indirect evidence. Hence, studies are needed that directly investigate this question using the same experimental design across age groups, as we aimed to do so in Study 4.

2.2. Does the consolidation of procedural knowledge follow an age-variant trajectory across the lifespan?

To overcome the limitation of previous studies comparing only two or a few age groups, in Study 4, we tested the 24-hour consolidation of statistical and general skill knowledge in a cross-sectional design with participants between the ages of 7 and 76 years. This lifespan design can help us draw a more sophisticated developmental curve on procedural memory consolidation. Considering statistical knowledge, both raw and standardized RT

data showed retained knowledge in all age groups, using both frequentist and Bayesian analyses. Regarding general skill knowledge, the analyses on raw RT data showed offline learning in all age groups, however, offline gain was significantly higher in the 7-8-year-olds than in the other age groups. Results on standardized RT data also suggested offline learning in all age groups, however, contrast to raw RT data, the offline gain was similar across the age groups. The uniform speed-up was corroborated by Bayesian analyses as well. Comparable retention of statistical knowledge across the age groups is in line with the results of Study 3 and with prior studies finding successful retention across various offline delays from childhood to old adulthood (e.g., Arciuli & Simpson, 2012; Hedenius et al., 2021; Hedenius et al., 2013; Kóbor et al., 2017; Nemeth, Janacsek, Király, et al., 2013), and with the findings of Study 4, direct evidence for age-invariant retention of statistical knowledge was provided.

3. Overall implications

To summarize the findings of the four studies, we can conclude that changes in the fronto-striatal networks might influence learning as shown by age-variant statistical learning across the lifespan (Janacsek et al., 2012) and by enhanced statistical learning in TS (Study 1). In contrast, retention of the acquired knowledge might not be influenced by changes in the fronto-striatal networks as consolidation is not altered in TS (Study 2) and age-invariant across the lifespan (Study 3 and 4). In the following section, I discuss the theoretical, methodological, and clinical relevance of the presented studies.

3.1. Theoretical implications of the presented studies

From a broader viewpoint, both Study 3 and 4 focused on the developmental trajectory of procedural memory consolidation. In Study 3, we indirectly investigated this question, whereas Study 4 directly examined the lifespan trajectory in a cross-sectional design. The results pointed to the same direction: the consolidation of statistical knowledge seems to be age-invariant across the lifespan.

As mentioned before, to the best of our knowledge, no theoretical model has been proposed to describe the lifespan trajectory of the consolidation of statistical knowledge, while the lifespan trajectory of statistical learning has been described with three different models (for a review, see Zwart et al., 2019). Janacsek et al. (2012) introduced the competition model to characterize the development of statistical learning. They proposed that during adolescence, a shift occurs from detecting raw probabilities to relying rather

on internal models, which results in a decreased statistical learning performance in adulthood than in childhood. On the neural level, detecting raw probabilities is related to the basal ganglia, particularly to the striatum, whereas internal models are rooted in the hippocampus and in the prefrontal cortex. According to the competition model, the early maturation of the striatum and the protracted development of the hippocampus and prefrontal cortex might explain the shift and the age variance in statistical learning (Janacsek et al., 2012).

Regarding the development of consolidation of statistical knowledge, no model has been proposed. Albouy et al. (2008) described a model for the consolidation of a related process, that is, motor sequence learning. They showed that in neurotypical young adults, the hippocampus and the striatum show an antagonistic relationship during the acquisition of a motor sequence and this dynamic is possibly mediated by the prefrontal cortex. During consolidation, however, the competitive dynamic is replaced by a cooperative dynamic. Relatedly, striatal activity seems to support the time-dependent maintenance in performance, that is, retention, while hippocampal activity seems to support the sleep-dependent enhancement in performance, that is, offline learning, at least in young adults (Albouy et al., 2008). Based on our behavioral results and the model of Albouy et al. (2008), tentative assumptions can be made for the neural network underlying the consolidation of statistical knowledge. We discovered time-dependent maintenance of statistical knowledge across the lifespan with comparable performance in all age groups, which suggests that this cognitive process might be more reliant on the striatum, which matures early in life. Consistently, the consolidation of statistical knowledge, at least when measured with the ASRT task, is independent of sleep (Nemeth, Janacsek, Londe, et al., 2010; Song et al., 2007b). It is important to note that this notion is highly speculative and further studies are warranted to corroborate it on the neural level.

An important theoretical aspect of learning and memory research in general is the dissociation of performance versus competence. This dissociation suggests that performance can fluctuate, and in a given moment, it might not always accurately reflect competence, that is, the underlying knowledge (Soderstrom & Bjork, 2015). As discussed several times throughout the dissertation, Janacsek et al. (2012) propose age-variant learning of statistical regularities across the lifespan. It is possible, however, that even though the results show a performance drop in statistical learning above the age of 12, competence of learning does not change. Janacsek et al. (2012) suggested that the decreased performance above the age of 12 might be related to the maturation of the

prefrontal cortex and to the emergence of internal models at the expense of extracting raw regularities from the environment. It might be possible that maturation of the prefrontal cortex-related networks only decreases the performance but not the competence of learning.

The dissociation of performance and competence can be tested by manipulating the temporal factors, for instance, the stimulus presentation rate. Kiss, Nemeth, and Janacsek (2022) manipulated the stimulus presentation rate in the ASRT task to test how the elapsed time between consecutive events influence the performance versus competence dissociation. Concerning reaction times, testing the acquired knowledge with faster presentation rates resulted in better learning, irrespective of the presentation rates during learning. This indicates that faster presentation rates can help with the expression of the acquired knowledge, at least in young adults. Employing the ASRT task, two studies also showed the dissociation of performance and competence by manipulating the task's instructions (i.e., focusing on accuracy or reaction time, Vékony et al., 2020; Vékony, Pleche, Pesthy, Janacsek, & Nemeth, 2022). It would be beneficial to test how temporal factors and instructions can influence the dissociation of performance and competence across the lifespan and how this might be related to the development of procedural memory.

3.2. Methodological implications of the presented studies

In clinical, developmental and lifespan studies involving the examination of RTs, taking into account the baseline RT differences across groups is highly important. Clinical groups sometimes show slower RTs and higher variance than their neurotypical peers (e.g., Janacsek, Borbély-Ipkovich, Nemeth, & Gonda, 2018) and several studies showed that children and older adults demonstrate slower RTs than young adults (Janacsek et al., 2012; Juhasz et al., 2019; Lukács & Kemény, 2015). The baseline differences in RT can influence and confound the learning performance: as clinical groups, children and older adults show slower baseline RTs, they have more room to improve in learning. In Study 1 and 2, participants with TS and typically developing peers showed similar baseline RTs, thus, standardization of RTs was not needed. In contrast, we found marked baseline RT differences in Study 3 and 4, where we tested the possibility of age variance in consolidation. Therefore, we analyzed standardized RT data in conjunction with raw RT data.

We believe that baseline RT differences must be considered; however, it is important to note that standardization of RT data has its own drawbacks as well. Both processing speed and variability are inherent aspects of development and aging. Slower processing speed and higher variability in learning can results in a better learning performance. Hence, during standardization, we risk the possibility of eliminating these fundamental aspects of learning. Moreover, transforming RT data can hinder interpretability, transformed results are often difficult to understand and explain. Relatedly, it is possible that across the lifespan, the relation between baseline RT and learning performance differs. For instance, Juhasz et al. (2019) aimed to explore the more room to improve concept and investigated how baseline RT differences between the ages of 7 and 85 influence statistical learning as well as general skill learning in the ASRT task. Baseline RTs and general skill learning were correlated within the age groups. They found significant positive correlations in children, in young adulthood and old adulthood, but not in adolescents and middle-aged participants, which suggests that this relationship is not universal across the lifespan. In contrast, no correlation was found between baseline RTs and statistical learning in any of the age groups, which suggests that the statistical learning measure used in the ASRT task is well-suited for group comparisons even when baseline RT differences across the groups are present (Juhasz et al., 2019). Nevertheless, in Study 4, we used both raw RT data and standardized RT scores. Results on general skill knowledge differed between the raw and standardized RT data, suggesting that the enhanced offline learning in the 7-8-year-old group might be due to the more room to improve concept. In contrast, results on the consolidation of statistical knowledge remained identical to the raw RT data across several standardization methods. In summary, based on the above methodological pitfalls, we suggest that both raw and standardized RT data should be reported in developmental and lifespan studies.

3.3. Clinical implications of the presented studies

Possibly enhanced learning and intact retention of procedural memories in TS has clinical as well as educational implications. As mentioned several times in the dissertation, procedural memory is related to cognitive, motor, and social skills and habits as well (Kaufman et al., 2010; Ullman, 2004). Employing a one-year retention in Study 2 helps to infer conclusions on real-world observations as learning a new skill or developing a habit take place over a longer time period than the usual time scale of lab experiments (typically hours or days). Thus, based on prior studies (Shephard et al., 2019; Takács et

al., 2018) and the results of Study 1 and 2, children with TS might be better in acquiring a new skill and they also might be successful in maintaining and remembering the learned skills. The results might also bear some clinical implications: the first-line treatments of tics are behavioral therapies, such as habit reversal training (Piacentini et al., 2010). During habit reversal training, when patients sense the urge to tic, they learn to carry out an oppositional action that is physically incompatible with the tic and more appropriate (e.g., flexing neck muscles instead of head twitching). Patients will eventually start performing the adequate action instead of the tic when feeling the urge, replacing the urge – tic association with the urge – adequate action association. It is plausible that, just as tics and procedural knowledge, the new urge – adequate action association leads to a stable memory representation (see more details in the discussion of Study 2).

4. Limitations and future directions

The presented studies are not without limitations. Considering Study 1 and 2, clinical samples are often heterogeneous across studies regarding age, medication, comorbid diagnoses and symptom severity, and the sample sizes are relatively low. The diversity of past research makes it difficult to draw firm conclusions on procedural hyperfunctioning in TS. Due to this inconsistency, it would be highly beneficial to conduct studies involving the same sample of TS participants and employ different kinds of statistical learning tasks. Relatedly, some well-established, reliable tasks measuring statistical learning have never been employed in a TS study, such as the embedded-pattern task. A large number of studies have employed variations of the embedded-pattern task using either spatial (Fiser & Aslin, 2001) or temporal visual or auditory streams (Turk-Browne, Isola, Scholl, & Treat, 2008). To comprehensively test how changes in the fronto-striatal networks lead to changes in procedural memory in TS, future studies should test the same large number of TS participants using not only variations of the embedded-pattern task but other task measuring different regularities within procedural memory as well.

The neural correlates of procedural memory have been thoroughly examined in previous studies. Concerning the ASRT task, both event-related potentials and neural oscillations have been tested before, at least in young adults (Kóbor et al., 2018; Tóth et al., 2017). Across the lifespan, the integration of brain networks undergo significant changes (e.g., Marek et al., 2018), therefore, event-related potentials and neural

oscillations related to procedural memory might also change during development. Employing the cued ASRT task, Kóbor et al. (2018) found that N2 amplitude modulation mirrors the changes on the behavioral level in young adults: probability-based regularities are acquired rapidly which was reflected in higher N2 amplitude for random high than random low trials and this difference remained constant across the task. N2 amplitude was also higher for pattern than random high trials and it increased with practice, reflecting the gradual acquisition of serial order-based regularities. Based on these results and the results of Study 1, we could expect that the speeded acquisition of probability-based regularities in TS would also be reflected on the neural level by enhanced N2 amplitudes. Moreover, past research have shown that in young adults, learning of statistical regularities is inversely correlated with the strength of connectivity in the theta band (Tóth et al., 2017). We could expect that as statistical learning might be enhanced in TS, this inverse relation might be stronger in TS than in typically developing peers.

Moreover, rewiring of procedural memory representations is an intriguing field of research and has been investigated in neurotypical adults (Horvath, Nemeth, & Janacsek, 2022; Szegedi-Hallgató et al., 2017). As mentioned before, according to past research, memory representations of procedural knowledge might be overstable in TS (Shephard et al., 2019; Takacs et al., 2021), which would lead to higher resistance to interference. This would make it more difficult to overwrite or rewire existing associations. A fruitful line of research has been investigating a related process, that is, stimulus-response (S-R) learning in TS. Beste and Münchau (2018) proposed that an enhanced cognitive function of acquiring S-R associations might explain tics in TS. Moreover, they hypothesized that the acquired associations are rigid, which would hinder the adaptation to contextual changes. Empirical findings corroborated this notion in adults (Kleimaker et al., 2020), but not in children with TS (Beste et al., 2021). Future research should examine the extent to which TS patients can rewire their procedural memory representations and how this affects their response to behavioral treatments.

Cross-sectional research design also has its limitations, past research has found a serious discrepancy between cross-sectional and longitudinal studies (Keresztes et al., 2022; Nyberg et al., 2010). For instance, one striking difference emerged in the field of hippocampal maturation: While cross-sectional findings implicated an increase in specific hippocampal subfields from middle childhood to early adolescence, longitudinal developmental trajectories suggested the opposite. Thus, investigating developmental differences using longitudinal designs instead of cross-sectional ones is warranted. One

of our ongoing studies aims to examine how statistical learning changes over time: we tested participants four times over an approximately 7-year time window, at 7, 8, 11 and 14 years of age. Preliminary results suggest that the result of Janacsek et al. (2012) showing decreased learning above the age of 12 has been replicated: over the span of seven years, participants showed a gradual decrease in learning in terms of raw reaction times. Hence, in this case, cross-sectional and longitudinal findings point toward the same direction. Nevertheless, in future studies, it would be beneficial to investigate whether age-invariance in procedural memory consolidation is supported by longitudinal findings as well.

5. Conclusions

The aim of the dissertation was to provide a deeper understanding of procedural memory by investigating this memory system in atypically and typically developing children and adolescents and across the lifespan. We investigated two regularities within the procedural memory system, that is, probability-based and serial-order regularities. Our results showed enhanced learning of probability-based regularities, whereas acquisition of serial-order regularities was impaired in Tourette Syndrome. The consolidation of probability-based regularities was robust in this neurodevelopmental disorder both over a short- and long-term offline delay. Similarly, consolidation of probability-based as well as serial-order regularities is resistant to forgetting over a one-year delay in typically developing children and adolescents. This is line with prior studies showing successful retention of statistical knowledge in neurotypical young adults (Kóbor et al., 2017; Romano et al., 2010), suggesting age-invariance in consolidation. The notion of age-invariant consolidation of statistical knowledge was corroborated by Study 4, where comparable retention was found from the age of 7 to 76 years. To sum up, based on past research and the results of the four studies presented in the dissertation, we can draw the following conclusions: changes in the fronto-striatal networks might influence the acquisition of procedural information, but not the retention of procedural knowledge.

Supplementary materials of Study 1 entitled

"Dissociation between two aspects of procedural learning in Tourette syndrome: Enhanced statistical and impaired sequence learning"

1. Supplementary data analyses on sample without comorbid diagnoses

In order to check whether comorbidities could confound the results or explain the procedural advantage reported in the manuscript, we have run the same analyses as described in the manuscript on the 17 children with TS without any comorbidities and their matched controls.

1.1. Supplementary Results

We ran a mixed design ANOVA on RT data across the four epochs. Statistical learning was quantified with a mixed design ANOVA with FREQUENCY (random high-frequency and random low-frequency triplets) and EPOCH (1-4) as within-subjects factors and GROUP (TS and TD) as a between-subjects factor. Sequence learning was also quantified with a mixed design ANOVA with ORDER (pattern high-frequency and random high-frequency triplets) and EPOCH (1-4) as within-subjects factors and GROUP (TS and TD) as a between-subjects factor. To test for post hoc pairwise comparisons, we used LSD (Least Significance Difference) tests.

As for **statistical learning**, the main effect of FREQUENCY was significant ($F(1, 32) = 56.11$, $p < .001$, $\eta^2_p = 0.637$), indicating that RTs were faster on random high-frequency triplets than on random low-frequency ones. The main effect of EPOCH was also significant ($F(3, 96) = 42.16$, $p < .001$, $\eta^2_p = 0.569$), meaning that over groups, participants became faster with practice on both triplets. Crucially, the FREQUENCY*EPOCH*GROUP interaction was significant ($F(3, 96) = 3.50$, $p = .018$, $\eta^2_p = 0.099$), meaning that the time course of statistical learning was different between the groups. Similarly to the results presented in the manuscript, follow-up analysis revealed a difference in the first epoch between the groups: The TS group showed higher learning than the TD group (TS: $M = 28.14$ ms, $SD = 28.95$ ms; TD: $M = -3.47$ ms, $SD = 29.55$ ms). There was no difference in the remaining epochs (all ps > .397). The main effect of GROUP and other interactions were not significant (all ps > .102).

As for **sequence learning**, the main effect of ORDER was significant ($F(1, 32) = 8.90$, $p = .005$, $\eta^2_p = 0.218$), meaning that participants showed faster RTs on the pattern

high-frequency triplets compared with the random high-frequency triplets. The main effect of EPOCH was significant as well ($F(3, 96) = 44.35$, $p < .001$, $\eta^2_p = 0.581$), suggesting that participants showed faster RTs with practice over both triplets. Similarly to the results in the manuscript, the ORDER*EPOCH interaction was significant ($F(3, 96) = 5.02$, $p = .032$, $\eta^2_p = 0.136$), meaning that the groups differed in the RT difference between the triplets. Follow-up analysis on the learning scores suggests that while the TD group learned to differentiate between the triplets, the TS group showed similar RTs on both triplets (TD: $M = 46.65$ ms, $SD = 71.27$ ms; TS: $M = 6.63$ ms, $SD = 18.51$ ms). The EPOCH*GROUP interaction was significant ($F(2.0, 65.9) = 3.77$, $p = .027$, $\eta^2_p = 0.106$), other main effects or interactions were not significant (all ps > .253).

To summarize, analyses without comorbidities showed *identical result as our original analyses,* indicating that the inclusion of participants with ADHD and OCD comorbidities in the TS group explains neither the procedural enhancement, nor the results of sequence learning.

Supplementary materials of Study 2 entitled

"Access to Procedural Memories After One Year: Evidence for Robust Memory Consolidation in Tourette Syndrome"

1. Analyses of the short- and long-term consolidation of serial-order knowledge

Analyses on performance in the Learning Phase have shown that the participants did not acquire the serial-order information (see Prerequisite of memory consolidation section in the Manuscript). Hence, the prerequisite of memory consolidation did not fulfill which calls into question the applicability of retention analyses concerning sequence learning. Nevertheless, in sake of completeness, we report the analyses on short- and long-term retention of serial-order knowledge here.

1.1. Short-term (five-hour) consolidation of serial-order knowledge

To test the 5-hour retention of serial-order knowledge, we run a mixed-design ANOVA on RT with GROUP (TS vs. TD) as between-subjects factor and ORDER (pattern vs. random high) and EPOCH (4 vs. 5) as within-subject factors. Overall, irrespective of epochs and group, participants were faster on pattern high (M = 386.37 ms) than on random high trials (M = 415.70 ms) (main effect of ORDER, $F(1, 36) = 5.58$, $p < 0.02$, $\eta^2_p = 0.13$). The ANOVA revealed retained serial-order memory after the 5-hour delay (non-significant ORDER × EPOCH interaction, $F(1, 36) = 2.20$, $p = 0.15$, $BF_{01} = 2.06$), memory scores were similar in the 4th (M = 22.22 ms) and 5th (M = 36.43 ms) epochs. The TS and TD groups showed comparable memory performance (non-significant GROUP × ORDER × EPOCH interaction, $F(1, 36) = 0.52$, $p = 0.48$, independent samples t-tests were conducted on the short-term offline change score, BF_{01} = 2.59, short-term offline change scores: M_{TS} = 21.11 ms, M_{TD} = 7.32 ms). Other main effects or interactions were also not significant (all ps > 0.070).

1.2. Long-term (one-year) consolidation of serial-order knowledge

To test 1-year retention of sequential knowledge, we conducted a mixed design ANOVA on RT with GROUP (TS vs. TD) as between-subjects factor and ORDER (pattern vs. random high) and EPOCH (6 vs. 7) as within-subject factors. Irrespective of group and epochs, participants showed faster RTs on pattern high (M = 370.03 ms) than on random high trials (M = 412.08 ms) (main effect of ORDER, $F(1, 36) = 5.34$, $p = 0.03$,

η^2_p = 0.13). The ANOVA revealed that, over groups, the memory scores did not change in the 1-year-long offline period (non-significant ORDER × EPOCH interaction, $F(1, 36) = 1.90$, $p = 0.18$, $BF_{01} = 2.41$), with similar memory scores in the 6th ($M = 58.65$ ms) and in the 7th ($M = 25.45$ ms) epochs. The TS and TD groups did not differ in retention (non-significant GROUP × ORDER × EPOCH interaction, $F(1, 36) = 1.20$, $p = 0.28$, independent samples t-tests were conducted on the long-term offline change score $BF_{01} = 1.98$, long-term offline change scores: $M_{TS} = -59.61$ ms, $M_{TD} = -6.79$ ms). Other main effects or interactions were not significant (all ps > 0.074).

Supplementary materials of Study 3 entitled

"Statistical and sequence learning lead to persistent memory in children after a one-year offline period"

1. Analysis of accuracy data

In the ASRT task, participants were provided with feedback about their performance, i.e., about their average RTs and accuracy, after each block. They were encouraged to keep accuracy above 92%, and the mean accuracy in the study was 92.29 % (*SD* = 3.38 %). High accuracy scores and relatively low variability in samples of neurotypical participants can hinder the detection of learning (Vékony et al., 2020); therefore, RTs could be considered a more appropriate measure of statistical and sequence learning. Based on this argument, we reported only the RT data in the Manuscript. Here, we report the analyses on accuracy values, which revealed similar results as the results on RT data.

1.1. Statistical analysis

Similarly to RT values (see Statistical analyses section of the Manuscript), prior developmental studies showed that age has a large effect on average accuracy (Janacsek et al., 2012; Juhasz et al., 2019; Zwart et al., 2019). To test this, we first calculated average accuracy over the 10 epochs (i.e., accuracy data was calculated on all trials, irrespective of trial types). We then correlated the average accuracy with age, which revealed a significant positive correlation ($r(68) = .32$, $p = .007$), showing that younger children were less accurate on the task. To control for the effect of average accuracy differences related to age on learning and consolidation of knowledge, we transformed the data in the following way. We divided each participants' raw accuracy values of each trial type and each epoch by their own average performance (i.e., average accuracy) in the first epoch of the task (for a similar approach, see Horvath et al., 2020; Nitsche et al., 2003). Participants' performance was around 1 at the beginning of the task and changed as the task progressed. Values above 1 indicated that responses were more accurate on a given trial type than the responses combined to all trial types (i.e., average accuracy) in the very first epoch of the task; and values below 1 meant that responses were less accurate on a given trial type compared to average accuracy in the first epoch. We conducted all analyses in the Supplementary Material on standardized accuracy.

Statistical learning score in the Learning Phase and memory scores in the Testing and Retesting Phases were quantified as the difference between random high and random low trial types in accuracy (accuracy for random high minus accuracy for random low trials). The learning and memory scores of sequence learning were calculated as the difference between pattern and random high trial types in accuracy (accuracy for pattern minus accuracy for random high trials). Higher scores indicate larger statistical or sequence learning/memory. To assess learning and the retention of knowledge, repeated measures ANOVAs and paired-samples t-tests were conducted on standardized accuracy data, separately for statistical and sequence learning. The Greenhouse-Geisser epsilon (ε) correction was used when necessary. Original *df* values and corrected *p* values (if applicable) are reported with partial eta-squared (η^2_p) as a measure of effect size. In conjunction with the frequentist analyses, we performed Bayesian paired-samples t-tests and calculated the Bayes Factor (BF) for the relevant comparisons as well.

1.2. Results

1.2.1. Prerequisite of memory consolidation

To assess memory consolidation, significant learning has to occur preceding the offline period. Therefore, as a first step, we conducted repeated-measures ANOVAs on the Learning Phase to confirm that significant learning has occurred concerning both statistical and sequence learning. ANOVAs were conducted on standardized accuracy separately for statistical and sequence learning.

Statistical learning during the Learning Phase wase tested with a two-way repeated-measures ANOVA with PROBABILITY (random high vs random low) and EPOCH (1-4) as within-subject factors. The ANOVA revealed significant statistical learning (main effect of PROBABILITY, $F(1, 69) = 33.65$, $p < .001$, $\eta^2_p = .33$). Post-hoc pairwise comparisons revealed higher accuracy on random high trials ($M = 1.003$) compared to random low trials ($M = 0.97$). Average accuracy (i.e., irrespective of trial types) did not change throughout the task (main effect of EPOCH, $F(3, 207) = 1.54$, $p = .22$). Statistical learning also did not change as the task progressed (PROBABILITY × EPOCH interaction, $F(3, 207) = 0.93$, $p = .43$, Fig. S4.1A).

To test **sequence learning** during the Learning Phase, similar two-way repeated-measures ANOVAs with ORDER (pattern vs random high) and EPOCH (1-4) as within-subject factors were conducted. The ANOVA revealed marginally significant learning (main effect of ORDER, $F(1, 69) = 3.39$, $p = .07$, $\eta^2_p = .05$), participants showed

marginally higher accuracy on pattern (M = 1.01) compared to random high trials (M = 1.003). Neither the average accuracy, nor the extent of sequence learning changed throughout the task (main effect of EPOCH, $F(3, 207) = 1.46$, $p = .24$ and ORDER × EPOCH interaction, $F(3, 207) = 2.39$, $p = .07$, respectively, Fig. S4.1B).

Furthermore, to investigate whether individual differences influence the learning on the task, we correlated statistical and sequence learning scores with working memory capacity, with percentage of perseverative errors on the WCST task, with socioeconomic status, and with total problem score on the SDQ. To control for multiple comparisons, we employed False Discovery Rate correction. None of the correlations reached significance (all ps > .128). We also rerun the ANOVAs on the sample without left-handed participants to control for handedness. The results were identical to the ones on the whole sample.

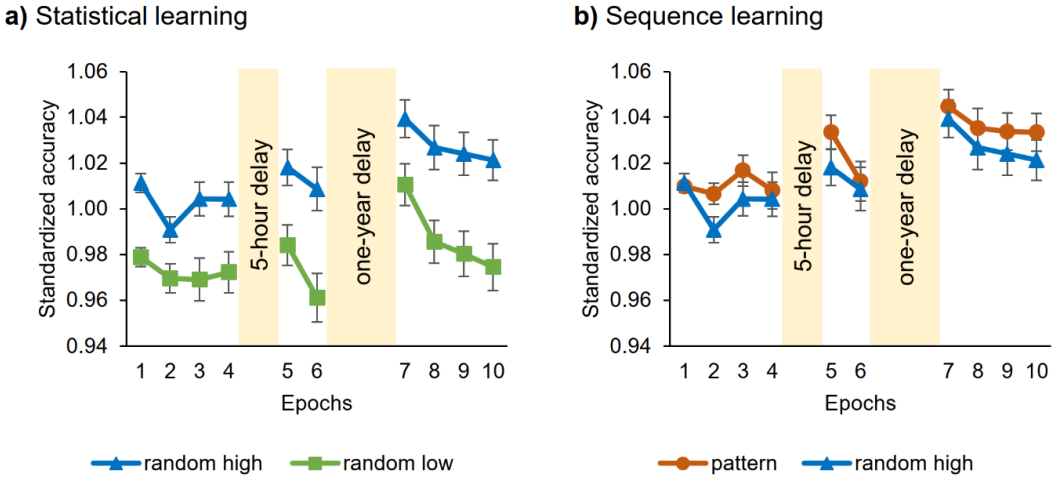

Figure S4.1. Temporal dynamics of (A) statistical and (B) sequence learning across epochs and sessions. Standardized accuracy values as a function of the epoch (1-10) and trial types (random high vs random low for statistical learning and pattern vs random high for sequence learning) are presented. Blue lines with triangle symbols indicate standardized accuracy values on the random high trials, green lines with square symbols indicate standardized accuracy values on the random low trials and orange lines with circle symbols indicate standardized accuracy values on the pattern trials. (A) Statistical learning is quantified by the gap between blue and green lines and (B) sequence learning is quantified by the gap between orange and blue lines. In both cases, greater gap between the lines represents better learning. Error bars denote standard error of mean.

1.2.2. Do children retain regularities after a one-year offline period?

To test one-year retention of **statistical knowledge**, we conducted a two-way repeated-measures ANOVA with PROBABILITY (random high vs random low) and EPOCH (6 vs. 7) as within-subject factors. Overall, irrespective of epochs, participants showed higher accuracy on random high (M = 1.02) compared to random low trials (M =

0.99) (main effect of PROBABILITY, $F(1, 69) = 49.53$, $p < .001$, $\eta^2_p = .42$). Average accuracy (i.e., irrespective of trial types) differed in the two epochs (main effect of EPOCH, $F(1, 69) = 30.10$, $p < .001$, $\eta^2_p = .30$), participants showed higher accuracy in the 7th epoch ($M = 1.03$) than in the 6th epoch ($M = 0.99$). The ANOVA showed a difference in memory scores between the Testing and Retesting Phases (significant PROBABILITY × EPOCH interaction, $F(1, 69) = 4.34$, $p = .04$, $\eta^2_p = .06$) Follow-up paired sample t-tests on the memory scores revealed that statistical memory underwent decrease over the one-year delay (6th epoch: $M = 0.0474$, 7th epoch: $M = 0.0288$; Fig. S4.2A), however, Bayesian analysis did not confirm this result ($BF_{01} = 1.003$). Furthermore, as the long-term delay had some variability in terms of weeks ($M_{delay} = 53.08$ weeks, $SD_{delay} = 2.39$ weeks, between 47.95 and 60.24 weeks), we examined whether it has any relation to the long-term memory performance. First, we calculated an offline change score for statistical knowledge by subtracting the standardized memory score in Epoch 6 from the standardized memory score in Epoch 7. This way, negative scores indicate forgetting, and positive scores indicate offline learning. Offline change score did not show correlation with the length of the long-term delay ($rs(68) = .129$, $p = .287$; $BF_{01} = 4.023$).

To investigate one-year retention of **serial-order knowledge**, we also ran a two-way repeated-measures ANOVA with ORDER (pattern vs random high) and EPOCH (6 vs. 7) as within-subject factors. Overall, irrespective of epoch, participants showed comparable accuracy on pattern and random high trials (main effect of PROBABILITY, $F(1, 69) = 1.71$, $p = .20$). Average accuracy (i.e., irrespective of trial types) differed in the two epochs (main effect of EPOCH, $F(1, 69) = 29.48$, $p < .001$, $\eta^2_p = .30$), participants showed higher accuracy in the 7th epoch ($M = 1.04$) than in the 6th epoch ($M = 1.01$). The ANOVA revealed evidence for persistent memory representations of serial-order knowledge (non-significant ORDER × EPOCH interaction, $F(1, 69) = 0.06$, $p = .81$, $BF_{01} = 7.404$). Follow-up paired sample t-tests on the memory scores showed comparable serial-order knowledge in the Testing and Retesting Phases (6th epoch: $M = 0.0035$, 7th epoch: $M = 0.0054$; Fig. S4.2B). Similarly to statistical knowledge, we also correlated the offline change score of serial-order knowledge and the length of the long-term delay. Offline change scores of serial-order knowledge did not correlate with the length of the delay ($rs(68) = -.190$, $p = .114$; $BF_{01} = 2.049$).

Moreover, similarly for the learning scores, to investigate whether individual differences influence the consolidation of statistical or serial-order knowledge, we

correlated the offline change scores with working memory capacity, with percentage of perseverative errors on the WCST task, with socioeconomic status and with total problem score on the SDQ. To control for multiple comparisons, we employed False Discovery Rate correction. None of the correlations reached significance (all *p*s > .766). We also rerun the ANOVAs on the sample without left-handed participants to control for handedness. The results were identical to the ones on the whole sample.

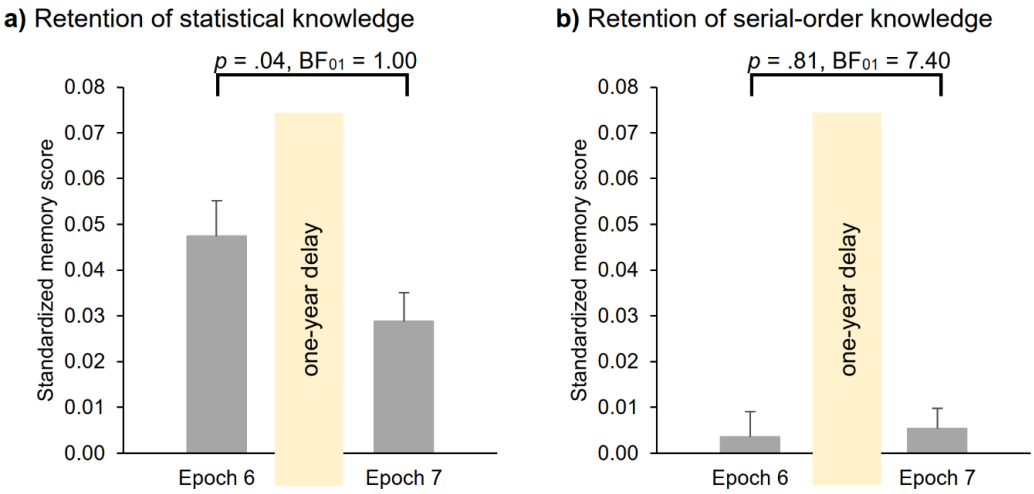

Figure S4.2. Retention of (A) statistical and (B) serial-order knowledge. Memory scores measured by standardized accuracy values for the last epoch of the Testing Phase (Epoch 6) and the first epoch of the Retesting Phase (Epoch 7). Error bars denote the standard error of mean.

1.2.3. Does age affect the one-year retention of statistical and serial-order regularities?

To check the possible association between age and retention, we conducted Pearson's correlation between the offline change scores and age. Regarding statistical knowledge, offline change scores did not show correlation with age ($r(68) = .01$, $p = .92$, $BF_{01} = 6.67$). Concerning serial-order knowledge, offline change scores in accuracy also did not correlate with age ($r(68) = .15$, $p = .21$, $BF_{01} = 3.08$).

Supplementary Material of Study 4 entitled

"Lifespan developmental invariance in memory consolidation: Evidence from procedural memory"

1. Bayesian model comparisons for consolidation of statistical and general skill knowledge

Table S5.1. Bayesian model comparisons for consolidation of statistical knowledge.

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.200	0.179	0.871	1.000	
Age group	0.200	0.717	10.122	0.250	0.480
Epoch + Age group	0.200	0.083	0.363	2.150	1.807
Epoch	0.200	0.021	0.085	8.583	1.343
Epoch + Age group + Epoch × Age group	0.200	3.382e -4	0.001	528.860	1.477

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

Table S5.2. Bayesian model comparisons for consolidation of general skill knowledge.

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.200	2.556e -46	1.022e -45	1.000	
Epoch + Age group	0.200	0.548	4.857	4.661e -46	1.032
Epoch + Age group + Epoch × Age group	0.200	0.452	3.295	5.659e -46	1.407
Age group	0.200	5.850e -18	2.340e -17	4.369e -29	0.328
Epoch	0.200	2.831e -29	1.132e -28	9.027e -18	0.964

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

2. Supplementary Figures of offline change of statistical and general skill knowledge

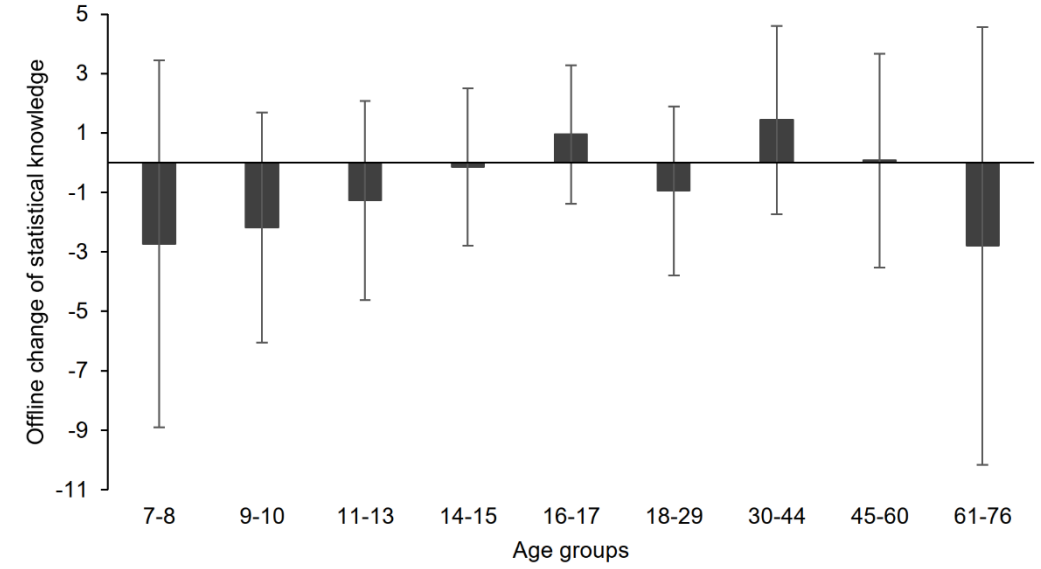

Figure S5.1. Offline change of statistical knowledge over the 24-hour delay across the age groups. Offline change scores were calculated by subtracting statistical knowledge values for the last epoch of the Learning Phase (Epoch 4) from those for the first epoch of the Testing Phase. Error bars denote the standard error of mean (SEM).

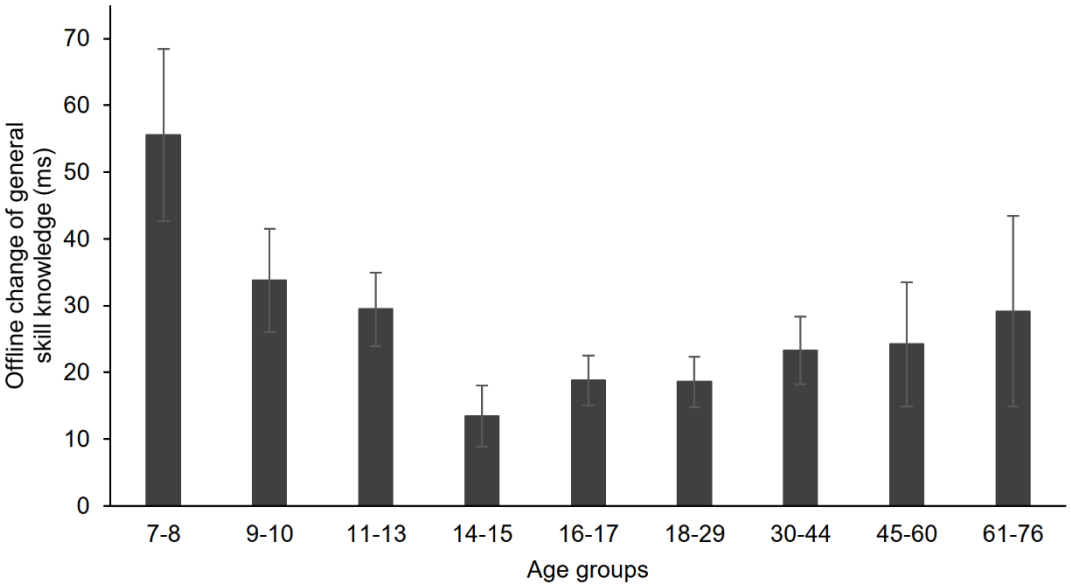

Figure S5.2. Offline change of general skill knowledge over the 24-hour delay across the age groups. Offline change scores were calculated by subtracting average RT values for the last epoch of the Learning Phase (Epoch 4) from those for the first epoch of the Testing Phase. Error bars denote the standard error of mean (SEM).

3. Results on learning

3.1. Are there age-related differences in statistical learning in terms of RTs?

To test whether the learning of statistical knowledge is age-variant, we submitted the statistical learning scores computed separately for the epochs of the Learning Phase (Epoch 1-4) to a mixed-design ANOVA with EPOCH (Epoch 1-4) as a within-subject factor and AGE GROUP as a between-subjects factor. The ANOVA revealed overall significant learning (main effect of INTERCEPT: $F(1, 246) = 225.24$, $p < .001$, $\eta_p^2 = 0.48$) and an overall increase of statistical learning score across the epochs (main effect of EPOCH: $F(3, 738) = 7.63$, $p < .001$, $\eta_p^2 = 0.03$). The trajectory of the increase was similar across the age groups (AGE GROUP x EPOCH interaction: $F(24, 738) = 0.88$, $p = 0.63$, $\eta_p^2 = 0.03$). Importantly, the ANOVA showed significant differences in overall learning across the age groups (main effect of AGE GROUP: $F(8, 246) = 2.69$, $p = .008$, $\eta_p^2 = 0.08$). Learning gradually decreased with age. In detail, the 7-8-year-olds showed better learning than participants over 14 ($p < .042$), the 9-10-year-olds showed better learning than participants over 16 ($p < .04$) and the 11-13-year-olds showed better learning than participants over 18 ($p < .045$). Over the age of 14, learning scores are comparable across the age groups ($p > .32$).

3.2. Are there age-related differences in general skill learning in terms of RTs?

To test whether general skill learning is age-variant, we submitted the median RTs of the epochs in the Learning Phase (Epoch 1-4) to a mixed-design ANOVA with EPOCH (Epoch 1-4) as a within-subject factor and AGE GROUP as a between-subject factor. The ANOVA showed that over groups, median RTs decreased as the task progressed (main effect of EPOCH: $F(3, 738) = 206.43$, $p < .001$, $\eta_p^2 = 0.46$) and median RTs differed significantly across age groups (main effect of AGE GROUP: $F(8, 246) = 27.90$, $p < .001$, $\eta_p^2 = 0.48$), revealing a U-shaped trajectory with highest RTs in the youngest and oldest age groups. Moreover, the trajectory of general skill learning differed across the age groups (AGE GROUP x EPOCH interaction: $F(24, 738) = 5.68$, $p < .001$, $\eta_p^2 = 0.16$). We calculated the amount of change in RTs from Epoch 1 to Epoch 4 by subtracting median RTs in Epoch 4 from median RTs in Epoch 1. This way, higher scores indicate a steeper decrease of reaction times, that is, better general skill learning. A follow-up one-way ANOVA on this score showed that 7-8-year-olds exhibited the greatest general skill

learning, which significantly differed from the other age groups' (ps < .01), except for the 61-76-year-old group (p = .34). The 9-10-year-olds showed smaller speed-up than the 7-8-year-olds, a comparable speed-up to the 11-13-year-olds and a higher speed-up than the other age groups between 14 and 60 years of age (p < .039). From 14 to 60 years of age, general skill learning was comparable across the age groups (p > .41). The 61-76-year-olds' general skill learning was significantly smaller than the other age groups (p < .006), expect for the 7-8-year-olds (p = .34) and 9-10-year-olds (p = .108).

4. Results on learning in terms of ratio scores

Due to the baseline RT differences across the age groups, we conducted additional ANOVA on ratio scores to test statistical learning while controlling for RT differences across the groups (for details on the standardization process, see Statistical analysis section of the manuscript). As the effect of standardization on general skill learning was comprehensively tested on this database in the study of Juhasz et al. (2019), here, we only report results on statistical learning in terms of ratio scores.

4.1. Are there age-related differences in statistical learning in terms of ratio scores?

Identically to the raw learning scores, we submitted the ratio scores to a mixed-design ANOVA with EPOCH (Epoch 1-4) as a within-subject factor and AGE GROUP as a between-subjects factor. The ANOVA revealed overall significant learning (main effect of INTERCEPT: $F(1, 246) = 278.07$, $p < .001$, $\eta_p^2 = 0.53$) and an overall increase of statistical learning score across the epochs (main effect of EPOCH: $F(3, 738) = 12.22$, $p < .001$, $\eta_p^2 = 0.05$). The trajectory of the increase was similar across the age groups (AGE GROUP x EPOCH interaction: $F(24, 738) = 0.93$, $p = 0.56$, $\eta_p^2 = 0.03$). The ANOVA also revealed significant differences in overall learning across the age groups (main effect of AGE GROUP: $F(8, 246) = 2.50$, $p = .01$, $\eta_p^2 = 0.08$), however, the differences were not identical to those of the analysis on raw statistical learning scores. Standardized statistical learning scores were comparable between the age of 7 and 29 ($p > .096$, $BF_{01} > 1.859$). The 30-44-year-old group exhibited decreased learning compared to 9-10-year-olds ($p = .05$, $BF_{01} = 1.142$) and 11-13-year-olds ($p = .02$, $BF_{01} = 0.485$), but there was no significant difference in the learning scores between the 30-44-year-old group and the age groups between 14 and 29 years (ps $> .087$, $BF_{01} > 1.438$). The 45-60-year-old and 61-76-year-old group showed decreased learning than almost all the age groups under the age of 15 ($p < .052$, $BF_{01} < 0.828$, except for the 7-8-year-olds vs. 61-76-year-olds, where $p = .057$, $BF_{01} = 0.654$).

5. Results on consolidation based on standardized reaction times

It is well-established that children and older adults respond with slower reaction times (RTs) overall (Janacsek et al., 2012; Juhasz et al., 2019). Hence, we conducted additional ANOVAs on standardized RTs to probe the retention of statistical and general skill knowledge while controlling for average RT differences across age groups. We standardized the data in two different ways: we calculated (1) ratio scores and (2) log-transformed RT data (for details on the standardization process, see Statistical analysis section of the manuscript). Here, we report the exact statistics of the analyses on standardized RTs.

5.1. Standardization with ratio scores

5.1.1. Do age groups differ in consolidation of statistical knowledge in terms of ratio scores?

To rule out the possibility of average RT differences among the age groups confounding our results, we tested the consolidation of statistical knowledge on ratio scores as well. We contrasted statistical learning scores computed from ratio scores for the last epoch of the Learning Phase (Epoch 4) with the learning scores computed for the first of epoch of the Testing Phase (Epoch 5) and submitted these scores to a mixed-design ANOVA with EPOCH (Epoch 4 vs Epoch 5) as a within-subject factor and AGE GROUP as a between-subjects factor. The ANOVA showed overall significant statistical knowledge (main effect of INTERCEPT: $F(1, 246) = 388.30$, $p < 0.001$, $\eta_p^2 = 0.61$) and significant differences in overall learning across age groups (main effect of AGE GROUP: $F(8, 246) = 4.70$, $p < 0.001$, $\eta_p^2 = 0.13$). Importantly, statistical knowledge appears to be retained over the 24-hour delay period with no significant change between the end of the Learning Phase and the Testing Phase (main effect of EPOCH: $F(1, 246) = 0.25$, $p = 0.62$, $\eta_p^2 = 0.001$). Moreover, no age group differences emerged in the retention of the statistical knowledge (non-significant EPOCH x AGE GROUP interaction: $F(8, 246) = 0.18$, $p = 0.99$, $\eta_p^2 = 0.006$): all age groups retained the acquired knowledge over the 24-hour delay period (all *p*s > .37). Bayesian mixed-design ANOVA on the standardized scores also supported the finding of the frequentist ANOVA by showing that the main effect of EPOCH and EPOCH x AGE GROUP interaction should be excluded from the model (Table S5.3 and S5.4). Thus, this analysis also suggests the successful retention of statistical knowledge over the 24-hour delay period in all age groups.

Table S5.3. Analysis of effects of Bayesian ANOVA for consolidation of standardized statistical knowledge in terms of ratio scores.

Effects	P(incl)	P(incl|data)	$BF_{exclusion}$
Epoch	0.400	0.099	9.081
Age group	0.400	0.998	0.002
Epoch × Age group	0.200	4.196e-4	235.891

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

Table S5.4. Bayesian model comparisons for standardized statistical knowledge in terms of ratio scores.

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.200	0.002	0.006	1.000	
Age group	0.200	0.899	35.556	0.002	0.494
Epoch + Age group	0.200	0.099	0.439	0.016	1.699
Epoch + Age group + Epoch × Age group	0.200	4.196e -4	0.002	3.690	0.965
Epoch	0.200	1.686e -4	6.747e -4	9.182	1.760

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

5.1.2. Do age groups differ in consolidation of general skill knowledge in terms of ratio scores?

Similarly to statistical knowledge, we tested consolidation of general skill knowledge (defined as overall RT changes) over the 24-hour delay period using ratio scores as well. We ran an ANOVA on ratio scores with EPOCH (Epoch 4 vs Epoch 5) as a within-subject factor and AGE GROUP as a between-subjects factor. The ANOVA revealed that RTs significantly decreased over the 24-hour delay (main effect of EPOCH: $F(1, 246) = 149.35$, $p < 0.001$, $\eta_p^2 = 0.38$), thus participants responded faster in the Testing Phase compared to the end of the Learning Phase (significant speed-up in all age groups: all *p*s

< 0.007). The amount of RT speed-up over the delay period was similar across age groups (non-significant EPOCH x AGE GROUP interaction: $F(8, 246) = 1.23$, $p = 0.28$, $\eta_p^2 = 0.04$; Fig. S5.3). Bayesian mixed-design ANOVA showed evidence for the inclusion of the main effect of EPOCH and the exclusion of the EPOCH x AGE GROUP interaction (Table S5.5 and S5.6), suggesting an overall speed-up over the delay period and a lack of differences in the amount of speed-up across age groups.

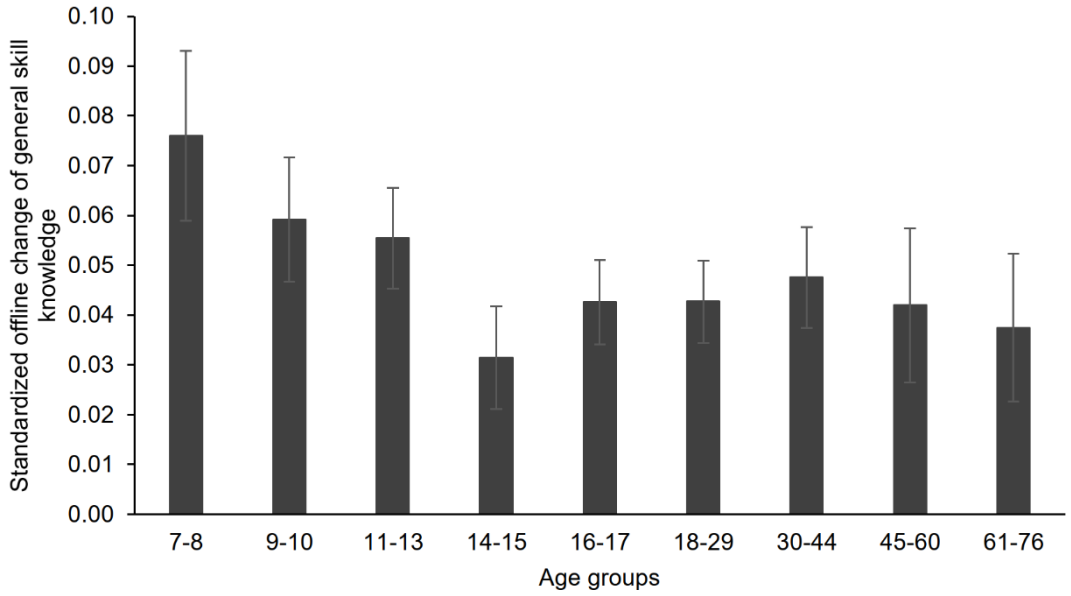

Figure S5.3. Standardized offline change of general skill knowledge over the 24-hour delay across the age groups. Standardized offline change scores were calculated by subtracting standardized average RT values for the last epoch of the Learning Phase (Epoch 4) from those for the first epoch of the Testing Phase. Error bars denote the SEM.

Table S5.5. Analysis of effects of Bayesian ANOVA for consolidation of standardized general skill knowledge in terms of ratio scores.

Effects	P(incl)	P(incl|data)	$BF_{exclusion}$
Epoch	0.400	0.945	1.728e-24
Age group	0.400	0.945	1.391e-6
Epoch × Age group	0.200	0.055	17.306

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

Table S5.6. Bayesian model comparisons for standardized general skill knowledge in terms of ratio scores.

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.200	2.010e -30	8.038e -30	1.000	
Epoch + Age group	0.200	0.945	69.222	2.126e -30	1.182
Epoch + Age group + Epoch × Age group	0.200	0.055	0.231	3.679e -29	2.024
Epoch	0.200	1.315e -6	5.262e -6	1.528e -24	2.009
Age group	0.200	1.634e -24	6.536e -24	1.230e -6	0.861

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

5.2. Standardization with log-transformation

For log-transformed RT data, we conducted identical ANOVAs as the ones presented in the manuscript and the ones for ratio scores presented in the supplementary materials, separately for statistical knowledge and general skill knowledge scores. Importantly, log-transformed standardization showed identical results as ratio score standardization both for statistical knowledge scores and general skill knowledge scores, both in frequentist and Bayesian ANOVAs. The results on statistical knowledge are presented in Table S5.7, S5.8 and S5.9; and the results on general skill knowledge are presented in Table S5.10, S5.11 and S5.12.

Table S5.7. The results of mixed-design ANOVA on the log-transformed statistical knowledge.

Effects	F	p	η_p^2
Intercept	380.35	< 0.001	0.61
Age group	4.44	< 0.001	0.13
Epoch	1.40	0.24	0.006
Epoch × Age group	0.80	0.60	0.025

Notes. Statistical knowledge is retained over the delay period in all age groups, with no age group differences in retention.

Table S5.8. Analysis of effects of Bayesian ANOVA for consolidation of log-transformed statistical knowledge.

Effects	P(incl)	P(incl|data)	$BF_{exclusion}$
Epoch	0.400	0.158	5.307
Age group	0.400	0.992	0.005
Epoch × Age group	0.200	0.003	46.144

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

Table S5.9. Bayesian model comparisons for log-transformed statistical knowledge.

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.200	0.004	0.015	1.000	
Age group	0.200	0.835	20.211	0.005	0.850
Epoch + Age group	0.200	0.157	0.747	0.024	1.953
Epoch + Age group + Epoch × Age group	0.200	0.003	0.014	1.110	1.550
Epoch	0.200	7.079e-4	0.003	5.345	1.865

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

Table S5.10. The results of mixed-design ANOVA on the log-transformed general skill knowledge.

Effects	F	p	η_p^2
Age group	31.53	< 0.001	0.51
Epoch	143.10	< 0.001	0.37
Epoch × Age group	1.33	0.23	0.04

Notes. All age groups showed offline learning (all *p*s < .050) and the speed-up over the delay period was similar across the age groups.

Table S5.11. Analysis of effects of Bayesian ANOVA for consolidation of log-transformed general skill knowledge.

Effects	P(incl)	P(incl|data)	$BF_{exclusion}$
Epoch	0.400	0.932	9.798e-24
Age group	0.400	0.932	6.645e-31
Epoch × Age group	0.200	0.068	13.638

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

Table S5.12. Bayesian model comparisons for log-transformed general skill knowledge.

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.200	4.944e-54	1.977e-53	1.000	
Epoch + Age group	0.200	0.932	54.552	5.306e-54	1.203
Epoch + Age group + Epoch × Age group	0.200	0.068	0.293	7.237e-53	1.700
Age group	0.200	9.129e-24	3.652e-23	5.415e-31	0.430
Epoch	0.200	6.191e-31	2.477e-30	7.985e-24	0.937

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

6. Testing age-related differences in consolidation by estimating future performance in the Testing Phase by extrapolation

6.1. Estimating statistical learning scores in the Testing Phase

Table S5.13. The results of the mixed-design ANOVA on the difference between the predicted and observed statistical learning scores.

Effects	F	p	η_p^2
Intercept	16.15	< 0.001	0.06
Block	1.56	0.18	0.006
Age group	1.53	0.15	0.05
Block × Age group	0.70	0.90	0.02

Table S5.14. Analysis of effects of the Bayesian ANOVA on the difference between the predicted and observed statistical learning scores.

Effects	**P(incl)**	**P(incl|data)**	**$BF_{exclusion}$**
Block	0.400	0.013	73.908
Age group	0.400	0.005	206.070
Block × Age group	0.200	1.270e-8	5234.761

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

Table S5.15. Bayesian model comparisons for the difference between the predicted and observed statistical learning scores.

Models	**P(M)**	**P(M|data)**	**BF_M**	**BF_{01}**	**error %**
Null model (incl. subject)	0.200	0.982	216.843	1.000	
Block	0.200	0.013	0.054	73.920	2.037
Age group	0.200	0.005	0.019	206.157	0.280
Block + Age group	0.200	6.647e-5	2.659e-4	14771.791	1.761
Block + Age group + Block × Age group	0.200	1.270e-8	5.079e-8	7.733e+7	0.687

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values

around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

6.2. Estimating general skill learning scores in the Testing Phase

Table S5.16. The results of the mixed-design ANOVA on the difference between the predicted and observed general skill learning scores.

Effects	F	p	η_p^2
Intercept	14.59	< 0.001	0.06
Block	6.84	< 0.001	0.03
Age group	0.47	0.88	0.02
Block × Age group	1.00	0.46	0.03

Table S5.17. Analysis of effects of the Bayesian ANOVA on the difference between the predicted and observed general skill learning scores.

Effects	**P(incl)**	**P(incl|data)**	**$BF_{exclusion}$**
Block	0.400	0.987	0.013
Age group	0.400	0.186	4.376
Block × Age group	0.200	6.190e-4	296.538

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

Table S5.18. Bayesian model comparisons for the difference between the predicted and observed general skill learning scores.

Models	**P(M)**	**P(M|data)**	**BF_M**	**BF_{01}**	**error %**
Null model (incl. subject)	0.200	0.010	0.041	1.000	
Block	0.200	0.803	16.347	0.013	0.975
Block + Age group	0.200	0.184	0.899	0.055	14.114
Age group	0.200	0.002	0.009	4.332	7.148
Block + Age group + Block × Age group	0.200	6.190e-4	0.002	16.272	66.534

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

7. Testing possible confounds influencing the consolidation of statistical and general skill knowledge

7.1. Block-level analysis on the consolidation of statistical knowledge

To test the possible confounding effect of averaging over the last five blocks of the Learning Phase and the first five blocks of the Testing Phase (Pan & Rickard, 2015), we contrasted performance in the last block of the Learning Phase (Block 20) and the first block is the Testing Phase (Block 21). We ran a mixed-design ANOVA on statistical learning scores with BLOCK (Block 20 vs Block 21) as a within-subject factor and AGE GROUP as a between-subjects factor. The analysis revealed that, on the group level, some degree of forgetting cannot be ruled out as the main effect of BLOCK was at the trend-level ($F(1, 246) = 3.54$, $p = 0.06$, $\eta_p^2 = 0.01$), learning scores were lower in Block 21 (M = 9.28 ms) than in Block 20 (M = 15.38 ms). Importantly, no significant age-related differences were detected in the retention of statistical knowledge (non-significant BLOCK x AGE GROUP interaction: $F(8, 246) = 1.67$, $p = 0.11$, $\eta_p^2 = 0.05$).

We also ran a Bayesian mixed-design ANOVA identical to the frequentist one. The ANOVA suggested that regarding the main effect of BLOCK, the data is not conclusive, the Bayes Factor was around 1, not supporting either the null or the alternative hypothesis (Table S5.19 and S5.20). Moreover, the lack of age-related differences in retention was supported by the Bayesian ANOVA (Table S5.19 and S5.20), corroborating the results of the frequentist ANOVA.

As noted in the main text, the analysis of block-wise data in the ASRT task should be interpreted carefully due to the relatively low number of trials. Statistical learning scores are calculated as difference scores between the high- and low-probability trials after excluding the first five random practice trials at the beginning of the block, erroneous responses as well as trills and repetitions from the 85 trials that are presented in a block. Hence, aggregated (mostly epoch-level) data has been used to characterize learning in the ASRT task since its inception because it enables to track the trajectory of learning while simultaneously decreasing the effect of noise in the learning scores to an acceptable level (J. H. Howard, Jr. & Howard, 1997; Song et al., 2007a, 2007b).

Table S5.19. Analysis of effects of the Bayesian ANOVA on statistical knowledge in the last block of the Learning Phase and in the first block of the Testing Phase.

Effects	P(incl)	P(incl|data)	$BF_{exclusion}$
Block	0.400	0.419	1.380
Age group	0.400	0.020	50.135
Block × Age group	0.200	0.002	3.762

Notes. The Effects column denotes the main effects and interaction. The P(incl) column indicates the prior inclusion probability and the P(incl|data) denotes the posterior inclusion probability. The $BF_{exclusion}$ column shows the exclusion Bayes Factors. $BF_{exclusion}$ values below 1 support the inclusion and values above 1 the exclusion of the given factor.

Table S5.20. Bayesian model comparisons for statistical knowledge in the last block of the Learning Phase and in the first block of the Testing Phase.

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.200	0.567	5.245	1.000	
Block	0.200	0.411	2.790	1.381	0.996
Age group	0.200	0.011	0.045	50.798	0.269
Block + Age group	0.200	0.008	0.034	67.997	1.591
Block + Age group + Block × Age group	0.200	0.002	0.009	255.803	1.778

Notes. All models include subject. The Models column denotes the predictors included in each model, the P(M) column the prior model probability, the P(M|data) column the posterior model probability, the BF_M column the posterior model odds, and the BF_{01} column the Bayes factors for each model compared to the null model. BF_{01} values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate substantial evidence and values larger than 10 indicate strong evidence for H_0. Values between 1 and 1/3 suggest anecdotal evidence, values between 1/3 and 1/10 indicate substantial evidence, and values below 1/10 indicate strong evidence for H_1. Values around 1 do not support either hypothesis. The error is an estimate of the numerical error in the computation of the Bayes factor.

7.2. Block-level analysis on the consolidation of general skill knowledge

To test whether practice-dependent changes in the Testing Phase influenced the offline learning of general skill knowledge over the 24-hour offline delay, we compared performance in the last block of the Learning Phase (Block 20) and the first block of the Testing Phase (Block 21). In detail, we contrasted general skill knowledge over the delay period with a mixed-design ANOVA on median RTs (i.e., RTs irrespective of the probabilities of events) with BLOCK (Block 20 vs Block 21) as a within-subject factor and AGE GROUP as a between-subject factor. The analysis showed that on the group level, median RTs significantly decreased over the 24-hour delay (main effect of BLOCK: $F(1, 246) = 11.00$, $p = 0.001$, $\eta_p^2 = 0.04$), participants responded faster in the first block of the Testing Phase compared to the last block of the Learning Phase. The

amount of speed-up was comparable across the age groups (as suggested by the non-significant BLOCK x AGE GROUP interaction: $F(8, 246) = 1.05$, $p = 0.40$, $\eta_p^2 = 0.03$). This suggests that practice-dependent changes did not influence offline learning of general skill knowledge over the offline delay.

To further investigate this effect, we conducted an additional ANOVA with PHASE (Learning Phase vs. Testing Phase) and BLOCK (1-5) as within-subject factors and AGE GROUP as a between-subjects factor. This way, we could compare how performance changed across Block 1-5 (first five blocks of the Learning Phase) and Block 21-25 (first five blocks of the Testing Phase). A similar increase in RTs in Block 1-5 and Block 21-25 would suggest that the offline learning showed over the offline delay could be explained by the additional practice in the Testing Phase rather than consolidation. The results are shown in Table S5.21.

Table S5.21. The results of the mixed-design ANOVA on general skill knowledge in the first five blocks of the Learning Phase and in the first five blocks of the Testing Phase.

Effects	F	p	η_p^2
Age group	28.51	< 0.001	0.48
Phase	554.79	< 0.001	0.69
Block	82.31	< 0.001	0.25
Age group × Phase	12.60	< 0.001	0.69
Age group × Block	2.48	< 0.001	0.08
Phase × Block	58.06	< 0.001	0.19
Age group × Phase × Block	2.26	0.001	0.07

The significant PHASE x BLOCK interaction suggests that the change in RTs across the blocks differed in the Learning and Testing Phases. Post-hoc analyses suggest that in the Learning Phase, there was a steeper increase in RTs ($M_{\text{Block 1}} = 625.56$ ms; $M_{\text{Block 5}} = 544.44$ ms, difference: 81.12 ms) than in the Testing Phase ($M_{\text{Block 21}} = 466.94$ ms, $M_{\text{Block 25}} = 454.49$ ms, difference: 12.45 ms). In the Learning Phase, all blocks differed from each other (all ps < 0.003) expect for Block 4 vs. 5 ($p = 0.27$). In the Testing Phase, RTs were slower in Block 21 compared to the remaining blocks (all ps < 0.001), but the remaining blocks did not differ from each other (all ps > 0.34). The steeper increase in RTs at the beginning of the Learning Phase than at the beginning of the Testing Phase

suggest that the observed offline learning of general skill knowledge over the 24-hour offline delay was not due to further practice-dependent changes in the Testing Phase.

The significant AGE GROUP x PHASE x BLOCK interaction suggests that the trajectory of performance change across the blocks in the Learning and Testing Phases differed in the age groups. To disentangle this effect, we ran two follow-up ANOVAs with BLOCK as a within-subject factor and AGE GROUP as a between-subject factor separately for the two phases. In the Learning Phase, the AGE GROUP x BLOCK interaction was significant: there was a steeper increase in RTs in the 7-8, 45-60 and 61-76-year-old groups compared to the other groups. In the Testing Phase, however, AGE GROUP x BLOCK interaction did not reach significance, suggesting that the trajectory of performance change was similar across the age groups.

7.3. The consolidation of general skill knowledge in terms of accuracy scores

To test whether offline learning in terms of reaction times could be influenced by decreased accuracy over the offline period, we ran a mixed-design ANOVA on mean accuracy scores (i.e., accuracy irrespective of the probabilities of events) with EPOCH (Epoch 4 vs Epoch 5) as a within-subject factor and AGE GROUP as a between-subject factor. The ANOVA revealed that mean accuracy scores significantly increased over the 24-hour delay (main effect of EPOCH: $F(1, 246) = 13.55$, $p < 0.001$, $\eta_p^2 = 0.05$). The amount of improvement was not uniform across the age groups (EPOCH x AGE GROUP interaction: $F(8, 246) = 4.12$, $p < .001$, $\eta_p^2 = 0.12$). Significant offline learning was only detectable in the 7-8-year-old group ($M_{Epoch\ 4} = 91.1\%$, $M_{Epoch\ 5} = 94.8\%$, $p < .001$) and in the 11-13-year-old group ($M_{Epoch\ 4} = 95.7\%$, $M_{Epoch\ 5} = 96.8\%$, $p = .045$). Mean accuracy scores were comparable over the delay in the other age groups (ps > .058). These results suggest that, in terms of accuracy scores, none of the age groups showed forgetting in general skills, therefore, offline learning in terms of reaction times cannot be explained by a decreased accuracy over the offline period.

References

Adi-Japha, E., Badir, R., Dorfberger, S., & Karni, A. (2014). A matter of time: rapid motor memory stabilization in childhood. *Developmental science, 17*(3), 424-433.

Albin, R., Koeppe, R. A., Bohnen, N. I., Nichols, T. E., Meyer, P., Wernette, K., . . . Frey, K. A. (2003). Increased ventral striatal monoaminergic innervation in Tourette syndrome. *Neurology, 61*(3), 310-315.

Albin, R., & Mink, J. W. (2006). Recent advances in Tourette syndrome research. *Trends in Neurosciences, 29*(3), 175-183.

Albouy, G., King, B. R., Maquet, P., & Doyon, J. (2013). Hippocampus and striatum: Dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. *Hippocampus, 23*(11), 985-1004.

Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., . . . Maquet, P. (2008). Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron, 58*(2), 261-272.

Alloway, T. P., & Alloway, R. G. (2013). Working memory across the lifespan: A cross-sectional approach. *Journal of Cognitive Psychology, 25*(1), 84-93.

APA. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is lasting and consistent over time. *Neuroscience Letters, 517*(2), 133-135.

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data. *Frontiers in Psychology, 3*, 324.

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

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

Avanzino, L., Martino, D., Bove, M., De Grandis, E., Tacchino, A., Pelosin, E., . . . Abbruzzese, G. J. M. D. (2011). Movement lateralization and bimanual coordination in children with Tourette syndrome. *26*(11), 2114-2118.

Barnes, K. A., Howard, J. H. J., Howard, D. V., Kenealy, L., & Vaidya, C. J. (2010). Two forms of implicit learning in childhood ADHD. *Developmental Neuropsychology, 35*(5), 494–505.

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.

Bauer, P. J., Dikmen, S. S., Heaton, R. K., Mungas, D., Slotkin, J., & Beaumont, J. L. (2013). III. NIH Toolbox Cognition Battery (CB): measuring episodic memory. *Monographs of the Society for Research in Child Development, 78*(4), 34-48.

Berg, E. A. (1948). A simple objective treatment for measuring flexibility in thinking. *Journal of General Psychology, 39*, 15-22.

Beste, C., Mückschel, M., Rauch, J., Bluschke, A., Takacs, A., Dilcher, R., . . . Li, S.-C. J. B. P. G. O. S. (2021). Distinct brain-oscillatory neuroanatomical architecture of perception-action integration in adolescents with Tourette syndrome. *1*(2), 123-134.

Beste, C., & Münchau, A. J. M. D. (2018). Tics and Tourette syndrome—surplus of actions rather than disorder? , *33*(2), 238-242.

Bloch, M. H., & Leckman, J. F. (2009). Clinical course of Tourette syndrome. *Journal of Psychosomatic Research, 67*(6), 497-501.

Bohlhalter, S., Goldfine, A., Matteson, S., Garraux, G., Hanakawa, T., Kansaku, K., . . . Hallett, M. (2006). Neural correlates of tic generation in Tourette syndrome: an event-related functional MRI study. *Brain, 129*(8), 2029-2037.

Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychologica, 128*(1), 33-44.

Bouyeure, A., & Noulhiane, M. (2020). Memory: Normative development of memory systems. In *Handbook of Clinical Neurology* (Vol. 173, pp. 201-213): Elsevier.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature, 382*(6588), 252-255. doi:10.1038/382252a0

Buse, J., Schoenefeld, K., Münchau, A., & Roessner, V. (2013). Neuromodulation in Tourette syndrome: dopamine and beyond. *Neuroscience and Biobehavioral Reviews, 37*(6), 1069-1084.

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.

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. (2001). Changes in executive control across the life span: examination of task-switching performance. *Developmental Psychology, 37*(5), 715.

Channon, S., Drury, H., Martinos, M., Robertson, M. M., Orth, M., & Crawford, S. J. N. (2009). Tourette's syndrome (TS): Inhibitory performance in adults with uncomplicated TS. *23*(3), 359.

Channon, S., Pratt, P., & Robertson, M. M. (2003). Executive function, memory, and learning in Tourette's syndrome. *Neuropsychology, 17*(2), 247-254.

Cleeremans, A., & Dienes, Z. (2008). Computational models of implicit learning. In R. Sun (Ed.), *The Cambridge Handbook of Computational Modeling* (pp. 396-421). Cambridge, UK: Cambridge University Press.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communication Disorders, 43*(1), 1-40.

Conceição, V. A., Dias, Â., Farinha, A. C., & Maia, T. V. (2017). Premonitory urges and tics in Tourette syndrome: computational mechanisms and neural correlates. *Current Opinion in Neurobiology, 46*, 187-199.

Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental Neuropsychology, 31*(1), 103-128.

Conway, C., & Christiansen, M. (2001). Sequential learning in non-human primates. *Trends in cognitive sciences, 5*(12), 539-546.

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.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*, 356–371.

Csábi, E., Benedek, P., Janacsek, K., Katona, G., & Nemeth, D. (2013). Sleep disorder in childhood impairs declarative but not nondeclarative forms of learning. *Journal of Clinical and Experimental Neuropsychology, 35*(7), 677-685.

Csabi, E., Benedek, P., Janacsek, K., Zavecz, Z., Katona, G., & Nemeth, D. (2015). Declarative and Non-declarative Memory Consolidation in Children with Sleep

Disorder. *Frontiers in Human Neuroscience, 9*, 709. doi:10.3389/fnhum.2015.00709

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. J. N. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *88*(1), 2-19.

Delorme, C., Salvador, A., Valabregue, R., Roze, E., Palminteri, S., Vidailhet, M., . . . Worbe, Y. (2016). Enhanced habit formation in Gilles de la Tourette syndrome. *Brain, 139*(2), 605-615.

Desmottes, L., Maillart, C., & Meulemans, T. (2017). Memory consolidation in children with specific language impairment: Delayed gains and susceptibility to interference in implicit sequence learning. *Journal of Clinical and Experimental Neuropsychology, 39*(3), 265-285.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science, 6*(3), 274-290. doi:10.1177/1745691611406920

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

Dikmen, S. S., Bauer, P. J., Weintraub, S., Mungas, D., Slotkin, J., Beaumont, J. L., . . . Heaton, R. K. (2014). Measuring episodic memory across the lifespan: NIH toolbox picture sequence memory test. *Journal of the International Neuropsychological Society, 20*(6), 611-619.

Dorfberger, S., Adi-Japha, E., & Karni, A. (2007). Reduced susceptibility to interference in the consolidation of motor memory before adolescence. *PloS One, 2*(2), 1-6.

Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., . . . Benali, H. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behavioral Brain Research, 199*(1), 61-75.

Dumay, N., Gaskell, M. G., & Feng, X. (2004). *A day in the life of a spoken word.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Dye, C. D., Walenski, M., Mostofsky, S. H., & Ullman, M. T. (2016). A verbal strength in children with Tourette syndrome? Evidence from a non-word repetition task. *Brain and Language, 160*, 61-70.

Eddy, C. M., & Cavanna, A. E. J. B. N. (2013). Altered social cognition in Tourette syndrome: nature and implications. *27*(1), 15-22.

Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. *Current Biology, 16*(13), 1290-1294.

Fanuel, L., Pleche, C., Vekony, T., Quentin, R., Janacsek, K., & Nemeth, D. (2020). The longer the better? General skill but not probabilistic learning improves with the duration of short rest periods. *bioRxiv*, 2020.2005.2012.090886. doi:10.1101/2020.05.12.090886

Fekete, R., Filep, O., Gyüre, T., Ujvári, K., Janacsek, K., & Németh, D. (2010). The examination of development of the working memory: New Hungarian standardised procedures. *Psychological Studies–Szeged*, 123-132.

Fischer, S., Wilhelm, I., & Born, J. (2007). Developmental Differences in Sleep's Role for Implicit Off-line Learning: Comparing Children with Adults. *Journal of Cognitive Neuroscience, 19*(2), 214-227.

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

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

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63*, 287-313.

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.

Frost, R. L., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition, 147*, 70-74.

Ganos, C., Garrido, A., Navalpotro-Gómez, I., Ricciardi, L., Martino, D., Edwards, M. J., . . . Bhatia, K. P. J. M. D. (2015). Premonitory urge to tic in Tourette's is associated with interoceptive awareness. *30*(9), 1198-1202.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*(2), 105-132.

Goetz, C. G., Pappert, E. J., Louis, E. D., Raman, R., & Leurgans, S. (1999). Advantages of a modified scoring method for the Rush Video-Based Tic Rating Scale. *Movement disorders: official journal of the Movement Disorder Society, 14*(3), 502-506.

Goodman, J., Marsh, R., Peterson, B. S., & Packard, M. G. (2014). Annual research review: the neurobehavioral development of multiple memory systems–implications for childhood and adolescent psychiatric disorders. *Journal of child psychology and psychiatry, 55*(6), 582-610.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry, 38*(5), 581-586.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 501-518.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (2002). Motor sequence learning with the nondominant left hand. *Experimental Brain Research, 146*, 369-378.

Gulya, M., Rossi-George, A., Hartshorn, K., Vieira, A., Rovee-Collier, C., Johnson, M. K., & Chalfonte, B. L. (2002). The development of explicit memory for basic perceptual features. *Journal of Experimental Child Psychology, 81*(3), 276-297.

Hallgato, 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.

Hedenius, M., Lum, J. A., & Bölte, S. (2020). Alterations of procedural memory consolidation in children with developmental dyslexia. *Neuropsychology*.

Hedenius, M., Lum, J. A., & Bölte, S. (2021). Alterations of procedural memory consolidation in children with developmental dyslexia. *Neuropsychology, 35*(2), 185.

Hedenius, M., Persson, J., Alm, P. A., Ullman, M. T., Howard, J. H., Jr., Howard, D. V., & Jennische, M. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in Developmental Disabilities, 34*(11), 3924-3935. doi:10.1016/j.ridd.2013.08.014

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., . . . Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. *Research in Developmental Disabilities, 32*(6), 2362-2375.

Henderson, L. M., Weighall, A. R., Brown, H., & Gareth Gaskell, M. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental science, 15*(5), 674-687.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*, 523-532.

Horvath, K., Nemeth, D., & Janacsek, K. (2022). Inhibitory control hinders habit change. *Scientific Reports*, *12*(1), 1-11.

Horvath, K., Torok, C., Pesthy, O., Nemeth, D., & Janacsek, K. (2020). Divided attention does not affect the acquisition and consolidation of transitional probabilities. *Scientific Reports*. doi:10.1038/s41598-020-79232-yda129685-49a7-438c-85e0-20e6384f29e9

Howard, D. V., Howard, J. H., Jr., Japikse, K., DiYanni, C., Thompson, A., & Somberg, R. (2004). Implicit sequence learning: effects of level of structure, adult age, and extended practice. *Psychology and Aging*, *19*(1), 79-92.

Howard, D. V., Howard, J. H., Jr., Japikse, K. C., DiYani, C., Thompson, A., & Somberg, R. (2004). Implicit sequence learning: Effects of level of structure, adult age, and extended practice. *Psychology and Aging*.

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

Ikier, S., Yang, L., & Hasher, L. (2008). Implicit proactive interference, age, and automatic versus controlled retrieval strategies. *Psychological Science*, *19*(5), 456-461.

Jackson, S. R., Sigurdsson, H. P., Dyke, K., Condon, M., & Jackson, G. M. J. J. o. N. (2021). The role of the cingulate cortex in the generation of motor tics and the experience of the premonitory urge-to-tic in Tourette syndrome. *15*(3), 340-362.

Janacsek, K., Borbély-Ipkovich, E., Nemeth, D., & Gonda, X. (2018). How can the depressed mind extract and remember predictive relationships of the environment? Evidence from implicit probabilistic sequence learning. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *81*, 17-24.

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.

Janacsek, K., Shattuck, K., Tagarelli, K., Lum, J., Turkeltaub, P., & Ullman, M. T. (2020). Sequence learning in the human brain: A functional neuroanatomical meta-analysis of serial reaction time studies. *Neuroimage*, *207*, 116387.

JASP, T. (2019). JASP Version 0.11.1 Computer software. In.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

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

Jung, J., Jackson, S. R., Nam, K., Hollis, C., & Jackson, G. M. J. J. o. N. (2015). Enhanced saccadic control in young people with T ourette syndrome despite slowed pro-saccades. *9*(2), 172-183.

Karmiloff-Smith, B. A. (1994). Beyond modularity: A developmental perspective on cognitive science. *European Journal of Disorders of Communication, 29*(1), 95-105.

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

Keresztes, A., Ngo, C. T., Lindenberger, U., Werkle-Bergner, M., & Newcombe, N. S. (2018). Hippocampal maturation drives memory from generalization to specificity. *Trends in cognitive sciences, 22*(8), 676-686.

Keresztes, A., Raffington, L., Bender, A. R., Bögl, K., Heim, C., & Shing, Y. L. J. D. C. N. (2022). Longitudinal developmental trajectories do not follow cross-sectional age associations in hippocampal subfield and memory development. *54*, 101085.

Keri, S., Szlobodnyik, C., Benedek, G., Janka, Z., & Gadoros, J. (2002). Probabilistic classification learning in Tourette syndrome. *Neuropsychologia, 40*(8), 1356-1362.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*(1), 171-184.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neuroscience Letters, 461*(2), 145-149.

Kiss, M., Nemeth, D., & Janacsek, K. J. C. (2022). Do temporal factors affect whether our performance accurately reflects our underlying knowledge? The effects of stimulus presentation rates on the performance versus competence dissociation. *157*, 65-80.

Kleimaker, M., Takacs, A., Conte, G., Onken, R., Verrel, J., Bäumer, T., . . . Beste, C. (2020). Increased perception-action binding in Tourette syndrome. *Brain*.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1*(2), 106-120.

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

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

Leckman, J. F., & Cohen, D. J. (1999). *Tourette's syndrome-tics, obsessions, compulsions: Developmental psychopathology and clinical care*: John Wiley & Sons Inc.

Leckman, J. F., Riddle, M. A., Hardin, M. T., Ort, S. I., Swartz, K. L., Stevenson, J., & Cohen, D. J. (1989). The Yale Global Tic Severity Scale: initial testing of a clinician-rated scale of tic severity. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*(4), 566-573.

Lieberman, M. D. (2000). Intuition: a social cognitive neuroscience approach. *Psychological Bulletin, 126*(1), 109-137.

Lukács, Á., & Kemény, F. (2015). Development of different forms of skill learning throughout the lifespan. *Cognitive Science, 39*(2), 383-404.

Maheu, M., Dehaene, S., & Meyniel, F. J. e. (2019). Brain signatures of a multiscale process of sequence learning in humans. *8*, e41541.

Maheu, M., Meyniel, F., & Dehaene, S. (2020). Rational arbitration between statistics and rules in human sequence learning. *bioRxiv*.

Maia, T. V., & Conceição, V. A. J. B. p. (2017). The roles of phasic and tonic dopamine in tic learning and expression. *82*(6), 401-412.

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience, 14*(2), 154-162.

Marek, S., Tervo-Clemmens, B., Klein, N., Foran, W., Ghuman, A. S., & Luna, B. J. P. b. (2018). Adolescent development of cortical oscillations: Power, phase, and support of cognitive maturation. *16*(11), e2004188.

Marsh, R., Alexander, G. M., Packard, M. G., Zhu, H., Wingard, J. C., Quackenbush, G., & Peterson, B. S. (2004). Habit learning in Tourette Syndrome, a translational neuroscience approach to developmental psychopathology. *Archives of General Psychiatry, 61*, 1259-1268.

McGaugh, J. L. (2000). Memory--A century of consolidation. *Science, 287*(5451), 248-251.

Meier, B., & Cock, J. (2014). Offline consolidation in implicit sequence learning. *Cortex, 57*, 156-166.

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology, 69*(3), 199.

Mink, J. W. (2001). Basal ganglia dysfunction in Tourette's syndrome: a new hypothesis. *Pediatric Neurology, 25*(3), 190-198.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1*, 31.

Morand-Beaulieu, S., Leclerc, J. B., Valois, P., Lavoie, M. E., O'Connor, K. P., & Gauthier, B. J. B. s. (2017). A review of the neuropsychological dimensions of Tourette syndrome. *7*(8), 106.

Mueller, S. C., Jackson, G. M., Ranu, D., Sophia, D., & Hollis, C. P. (2006). Enhanced cognitive control in young people with Tourette's syndrome. *Current Biology, 16*(6), 570-573.

Mueller, S. T., & Piper, B. J. J. J. o. n. m. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *222*, 250-259.

Nemeth, D., & Janacsek, K. (2011). The dynamics of implicit skill consolidation in young and elderly adults. *Journal of Gerontology Psychological Science, 66*(1), 15-22.

Nemeth, D., Janacsek, K., Balogh, V., Londe, Z., Mingesz, R., Fazekas, M., . . . Vetro, A. (2010). Learning in Autism: Implicitly Superb. *PloS One, 5*(7), e11731.

Nemeth, D., Janacsek, K., Csifcsak, G., Szvoboda, G., Howard, J. H., Jr., & Howard, D. V. (2011). Interference between sentence processing and probabilistic implicit sequence learning. *PloS One, 8*(6(3)), e17577.

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*. doi:10.3389/fncom.2013.00147

Nemeth, D., Janacsek, K., Király, K., Londe, Z., Németh, K., Fazekas, K., . . . Csányi, A. (2013). Probabilistic sequence learning in mild cognitive impairment. *Frontiers in Human Neuroscience, 7*, 318. doi:10.3389/fnhum.2013.00318

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

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

Nitsche, M. A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., & Tergau, F. (2003). Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. *Journal of Cognitive Neuroscience, 15*(4), 619-626.

Nyberg, L., Salami, A., Andersson, M., Eriksson, J., Kalpouzos, G., Kauppi, K., . . . Nilsson, L.-G. J. P. o. t. N. A. o. S. (2010). Longitudinal evidence for diminished frontal cortex function in aging. *107*(52), 22682-22686.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Palminteri, S., Lebreton, M., Worbe, Y., Hartmann, A., Lehéricy, S., Vidailhet, M., . . . Pessiglione, M. (2011). Dopamine-dependent reinforcement of motor skill learning: evidence from Gilles de la Tourette syndrome. *Brain, 134*(8), 2287-2301.

Pan, S. C., & Rickard, T. C. (2015). Sleep and motor learning: is there room for consolidation? *Psychological Bulletin, 141*(4), 812.

Peterson, B. (1999). Neuroanatomical circuitry. In J. F. Leckman & D. J. Cohen (Eds.), *Tourette's syndrome: Tics, obsessions, compulsions. Developmental psychopathology and clinical care* (pp. 230-259). Hoboken, New Jersey: John Wiley & Sons.

Peterson, B., Skudlarski, P., Anderson, A., Zhang, H., Gatenby, J., Lacadie, C., . . . Gore, J. (1998). A functional magnetic resonance imaging study of tic suppression in Tourette syndrome. *Archives of General Psychiatry, 55*(4), p326-333.

Peterson, B., Thomas, P., Kane, M., Scahill, L., Zhang, H., Bronen, R., . . . Staib, L. (2003). Basal Ganglia volumes in patients with Gilles de la Tourette syndrome. *Archives of General Psychiatry, 60*(4), p415-424.

Petruo, V., Bodmer, B., Bluschke, A., Münchau, A., Roessner, V., & Beste, C. (2020). Comprehensive Behavioral Intervention for Tics reduces perception-action binding during inhibitory control in Gilles de la Tourette syndrome. *Scientific Reports, 10*(1), 1-8.

Petruo, V., Bodmer, B., Brandt, V. C., Baumung, L., Roessner, V., Münchau, A., . . . Psychiatry. (2019). Altered perception-action binding modulates inhibitory control in Gilles de la Tourette syndrome. *60*(9), 953-962.

Piacentini, J., & Chang, S. (2005). Habit reversal training for tic disorders in children and adolescents. *Behavior Modification, 29*(6), 803-822.

Piacentini, J., Woods, D. W., Scahill, L., Wilhelm, S., Peterson, A. L., Chang, S., . . . Levi-Pearl, S. (2010). Behavior therapy for children with Tourette disorder: a randomized controlled trial. *JAMA, 303*(19), 1929-1937.

Picard, L., Cousin, S., Guillery-Girard, B., Eustache, F., & Piolino, P. (2012). How do the different components of episodic memory develop? Role of executive functions and short-term feature-binding abilities. *Child Development, 83*(3), 1037-1050.

Piper, B. J., Mueller, S. T., Geerken, A. R., Dixon, K. L., Kroliczak, G., Olsen, R. H., & Miller, J. K. (2015). Reliability and validity of neurobehavioral function on the Psychology Experimental Building Language test battery in young adults. *PeerJ, 3*, e1460.

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.

Prehn-Kristensen, A., Göder, R., Chirobeja, S., Breßmann, I., Ferstl, R., & Baving, L. (2009). Sleep in children enhances preferentially emotional declarative but not procedural memories. *Journal of Experimental Child Psychology, 104*(1), 132-139.

Press, D. Z., Casement, M. D., Pascual-Leone, A., & Robertson, E. M. (2005). The time course of off-line motor sequence learning. *Cognitive Brain Research, 25*(1), 375-378.

Quentin, R., Fanuel, L., Kiss, M., Vernet, M., Vékony, T., Janacsek, K., . . . Nemeth, D. (2021). Statistical learning occurs during practice while high-order rule learning during rest period. *Npj Science Of Learning*, in press.

Rathbone, C. J., Moulin, C. J., & Conway, M. A. (2008). Self-centered memories: The reminiscence bump and the self. *Memory & Cognition, 36*(8), 1403-1414.

Reber. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious* (Vol. 19). New York: Oxford University Press.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855-863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219-235. doi:10.1037/0096-3445.118.3.219

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

Rickard, T. C. (2007). Forgetting and learning potentiation: dual consequences of between-session delays in cognitive skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 297.

Rickard, T. C., & Pan, S. C. (2017). Time for considering the possibility that sleep plays no unique role in motor memory consolidation: Reply to Adi-Japha and Karni (2016).

Robertson, E. M. (2009). From creation to consolidation: A novel framework for memory processing. *PLoS Biology, 7*(1), e1000019.

Robertson, E. M., Pascual-Leone, A., & Miall, R. C. (2004). Current concepts in procedural consolidation. *Nature Reviews Neuroscience, 5*(7), 576-582.

Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2004). Awareness modifies the skill-learning benefits of sleep. *Current Biology, 14*(3), 208-212. doi:10.1016/j.cub.2004.01.027

Robertson, M. M. (2015). A personal 35 year perspective on Gilles de la Tourette syndrome: prevalence, phenomenology, comorbidities, and coexistent psychopathologies. *The Lancet Psychiatry, 2*(1), 68-87.

Robertson, M. M., Eapen, V., Singer, H. S., Martino, D., Scharf, J. M., Paschou, P., . . . Mathews, C. A. (2017). Gilles de la Tourette syndrome. *Nature reviews Disease primers, 3*(1), 1-20.

Roessner, V., Overlack, S., Schmidt-Samoa, C., Baudewig, J., Dechent, P., Rothenberger, A., & Helms, G. (2011). Increased putamen and callosal motor subregion in treatment-naïve boys with Tourette syndrome indicates changes in the

bihemispheric motor network. *Journal of child psychology and psychiatry, 52*(3), 306-314.

Roig, M., Ritterband-Rosenbaum, A., Lundbye-Jensen, J., & Nielsen, J. B. J. N. o. A. (2014). Aging increases the susceptibility to motor memory interference and reduces off-line gains in motor skill learning. *35*(8), 1892-1900.

Romano Bergstrom, J. C., Howard, J. H., Jr., & Howard, D. V. (2012). Enhanced Implicit Sequence Learning in College-age Video Game Players and Musicians. *Applied Cognitive Psychology, 26*(1), 91-96.

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

Roth, R. M., Baribeau, J., Milovan, D., O'Connor, K., & Todorov, C. (2004). Procedural and declarative memory in obsessive-compulsive disorder. *Journal of the International Neuropsychological Society, 10*(5), 647-654.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review, 16*(2), 225-237. doi:doi:10.3758/pbr.16.2.225

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

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

Schapiro, A. C., Reid, A. G., Morgan, A., Manoach, D. S., Verfaellie, M., & Stickgold, R. (2019). The hippocampus is necessary for the consolidation of a task that does not require the hippocampus for initial learning. *Hippocampus, 29*(11), 1091-1100.

Shephard, E., Groom, M. J., & Jackson, G. M. (2019). Implicit sequence learning in young people with Tourette syndrome with and without co-occurring attention-deficit/hyperactivity disorder. *Journal of Neuropsychology, 13*(3), 529-549.

Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass, 14*(3), e12365.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Phil. Trans. R. Soc. B, 372*(1711), 20160059.

Simor, P., Zavecz, Z., Horvath, K., Elteto, N., Török, C., Pesthy, O., . . . Nemeth, D. (2019). Deconstructing procedural memory: Different learning trajectories and consolidation of sequence and statistical learning. *Frontiers in Psychology, 9*, 2708.

Singer, H. S. (2013). The neurochemistry of Tourette syndrome.

Smalle, E. H., Page, M. P., Duyck, W., Edwards, M., & Szmalec, A. (2018). Children retain implicitly learned phonological sequences better than adults: A longitudinal study. *Developmental science, 21*(5), e12634.

Smith, F. R., Gaskell, M. G., Weighall, A. R., Warmington, M., Reid, A. M., & Henderson, L. M. (2018). Consolidation of vocabulary is associated with sleep in typically developing children, but not in children with dyslexia. *Developmental science, 21*(5), e12639.

Soderstrom, N. C., & Bjork, R. A. J. P. o. P. S. (2015). Learning versus performance: An integrative review. *10*(2), 176-199.

Soetens, E., Melis, A., & Notebaert, W. (2004). Sequence learning and sequential effects. *Psychological Research, 69*(1), 124-137.

Song, S., Howard, J. H., Jr., & Howard, D. V. (2007a). Implicit probabilistic sequence learning is independent of explicit awareness. *Learning and Memory, 14*, 167–176.

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

Squire, L. R. (1994). Declarative and Nondeclarative Memory: Multiple Brain Systems Supporting Learning and Memory. In D. L. Schacter & E. Tulving (Eds.), *Memory Systems 1994* (pp. 407). Cambridge: The MIT Press.

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience, 34*, 259-288.

Stark-Inbar, A., Raza, M., Taylor, J. A., & Ivry, R. B. (2016). Individual differences in implicit motor learning: task specificity in sensorimotor adaptation and sequence learning. *Journal of Neurophysiology, 117*(1), 412-428.

Stebbins, G. T., Singh, J., Weiner, J., Wilson, R. S., Goetz, C., & Gabrieli, J. D. E. (1995). Selective impairments of memory functioning in unmedicated adults with Gilles de la Tourette's syndrome. *Neuropsychology, 9*(3), 329-337.

Stern, E., Silbersweig, D. A., Chee, K. Y., Holmes, A., Robertson, M. M., Trimble, M., . . . Dolan, R. J. (2000). A functional neuroanatomy of tics in Tourette syndrome. *Archives of General Psychiatry, 57*(8), 741-748.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford: Oxford University Press.

Szegedi-Hallgató, E., Janacsek, K., & Nemeth, D. J. P. o. (2019). Different levels of statistical learning-hidden potentials of sequence learning tasks. *14*(9), e0221966.

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

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

Takacs, A., Kobor, A., Kardos, Z., Janacsek, K., Horvath, K., Beste, C., & Nemeth, D. (2020). Neurophysiological coding of statistical and deterministic rule information. *bioRxiv*.

Takács, Á., Kóbor, A., Kardos, Z., Janacsek, K., Horváth, K., Beste, C., & Nemeth, D. (2021). Neurophysiological and functional neuroanatomical coding of statistical and deterministic rule information during sequence learning. *Human Brain Mapping*.

Takacs, A., Münchau, A., Nemeth, D., Roessner, V., & Beste, C. (2021). Lower-level associations in Gilles de la Tourette syndrome: convergence between hyperbinding of stimulus and response features and procedural hyperfunctioning theories. *European Journal of Neuroscience*.

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

Tanczos, T., Janacsek, K., & Nemeth, D. (2013a). [Verbal fluency tasks I. Investigation of the Hungarian version of the letter fluency task between 5 and 89 years of age]. *Psychiatria Hungarica: A Magyar Pszichiátriai Társaság Tudományos Folyóirata, 29*(2), 158-180.

Tanczos, T., Janacsek, K., & Nemeth, D. (2013b). [Verbal fluency tasks II. Investigation of the Hungarian version of the semantic fluency task between 5 and 89 years of age]. *Psychiatria Hungarica: A Magyar Pszichiátriai Társaság Tudományos Folyóirata, 29*(2), 181-207.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin, 139*(4), 792.

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of Developmental Differences in Implicit Sequence Learning: An fMRI Study of Children and Adults. *Journal of Cognitive Neuroscience, 16*(8), 1339-1351.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*(1), 1-42.

Tóth-Fáber, E., Janacsek, K., & Németh, D. (2021). Statistical and sequence learning lead to persistent memory in children after a one-year offline period. *Scientific Reports*, in press.

Tóth-Fáber, E., Tárnok, Z., Janacsek, K., Kóbor, A., Nagy, P., Farkas, B. C., . . . Nemeth, D. (2021). Dissociation between two aspects of procedural learning in Tourette syndrome: Enhanced statistical and impaired sequence learning. *Child Neuropsychology*, 1-23.

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*.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Turi, E., Tóth, I., & Gervai, J. (2011). Further examination of the Strengths and Difficulties Questionnaire (SDQ-Magy) in a community sample of young adolescents. *Psychiatria Hungarica: A Magyar Pszichiátriai Társaság Tudományos Folyóirata, 26*(6), 415-426.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(2), 399.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience, 30*(33), 11177-11187. doi:10.1523/jneurosci.0858-10.2010

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1-2), 231-270.

Ullman, M. T. (2016). The declarative/procedural model: a neurobiological model of language learning, knowledge, and use. In *Neurobiology of language* (pp. 953-968): Elsevier.

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual Review of Psychology, 71*.

Urbain, C., Houyoux, E., Albouy, G., & Peigneux, P. J. J. o. s. r. (2014). Consolidation through the looking-glass: sleep-dependent proactive interference on visuomotor adaptation in children. *23*(1), 44-52.

Vékony, T., Ambrus, G. G., Janacsek, K., & Nemeth, D. (2021). Cautious or causal? Key implicit sequence learning paradigms should not be overlooked when assessing the role of DLPFC (Commentary on Prutean et al.). *Cortex*.

Vékony, T., Marossy, H., Must, A., Vécsei, L., Janacsek, K., & Nemeth, D. (2020). Speed or accuracy instructions during skill learning do not affect the acquired knowledge. *Cerebral Cortex Communications, 1*(1), tgaa041.

Vékony, T., Pleche, C., Pesthy, O., Janacsek, K., & Nemeth, D. J. n. S. o. L. (2022). Speed and accuracy instructions affect two aspects of skill learning differently. *7*(1), 27.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychon Bull Rev, 14*(5), 779-804. doi:10.3758/BF03194105

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (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. doi:10.1037/a0022790

Walenski, M., Mostofsky, S. H., & Ullman, M. T. (2007). Speeded processing of grammar and tool knowledge in Tourette's syndrome. *Neuropsychologia, 45*, 2447–2460.

Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *Behavioral and Brain Sciences, 28*(1), 51-104.

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425*(6958), 616-620.

Wang, Z., Maia, T. V., Marsh, R., Colibazzi, T., Gerber, A., & Peterson, B. S. (2011). The neural circuits that generate tics in Tourette's syndrome. *American Journal of Psychiatry, 168*(12), 1326-1337.

Wilhelm, I., Metzkow-Mészàros, M., Knapp, S., & Born, J. (2012). Sleep-dependent consolidation of procedural motor memories in children and adults: The pre-sleep level of performance matters. *Developmental science, 15*(4), 506-515.

Worbe, Y., Gerardin, E., Hartmann, A., Valabregue, R., Chupin, M., Tremblay, L., . . . Lehericy, S. (2010). Distinct structural changes underpin clinical phenotypes in patients with Gilles de la Tourette syndrome. *Brain, 133*, 3649-3660.

Worbe, Y., Malherbe, C., Hartmann, A., Pélégrini-Issac, M., Messé, A., Vidailhet, M., . . . Benali, H. J. B. (2012). Functional immaturity of cortico-basal ganglia networks in Gilles de la Tourette syndrome. *135*(6), 1937-1946.

Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Valabregue, R., Poupon, F., Guevara, P., . . . Lehericy, S. (2015). Altered structural connectivity of cortico-striato-pallido-thalamic networks in Gilles de la Tourette syndrome. *Brain, 138*(2), 472-482.

Yaniv, A., Benaroya-Milshtein, N., Steinberg, T., Ruhrrman, D., Apter, A., & Lavidor, M. J. R. i. d. d. (2017). Specific executive control impairments in Tourette syndrome: The role of response inhibition. *61*, 1-10.

Zavecz, Z., Janacsek, K., Simor, P., Cohen, M. X., & Nemeth, D. (2020). Similarity of brain activity patterns during learning and subsequent resting state predicts memory consolidation. *bioRxiv*.

Zelazo, P. D., Craik, F. I. M., & Booth, L. (2004). Executive function across the life span. *Acta Psychologica, 115*, 167-183.

Zwart, F. S., Vissers, C. T. W., Kessels, R. P., & Maes, J. H. (2019). Procedural learning across the lifespan: A systematic review with implications for atypical development. *Journal of Neuropsychology, 13*(2), 149-182.