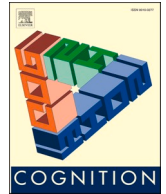


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Cognition

journal homepage: www.elsevier.com/locate/cognit

Prompting teaching modulates children's encoding of novel information by facilitating higher-level structure learning and hindering lower-level statistical learning[☆]

Hanna Marno^{a,*}, Róbert Danyi^a, Teodóra Vékony^b, Karolina Janacsek^{c,d,e,1},
Dezső Németh^{b,d,e,1}

^a Department of Cognitive Science, Central European University, Budapest 1051, Hungary

^b Lyon Neuroscience Research Center (CRNL), INSERM, CNRS, Université Claude Bernard Lyon 1, Centre Hospitalier Le Vinatier, Bâtiment 462 - Neurocampus 95 boulevard Pinel, 69675 Bron, France

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

^d Institute of Psychology, ELTE Eötvös Loránd University, Izabella utca 46, 1064 Budapest, Hungary

^e Brain, Memory and Language Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Magyar tudósok körútja 2, H-1117 Budapest, Hungary

A B S T R A C T

Young children are not only prepared to learn from teaching, but they also start to spontaneously teach others, indicating that teaching is a natural instinct of the humankind. During childhood, teaching seems to precede the emergence of several cognitive abilities, so the question arises: how does teaching affect the development of later emerging cognitive skills? Since teaching requires explicit, accessible representations of the knowledge of the teacher, we hypothesized that the motivation to teach might impact the way children encode novel information, by biasing them towards a model-based encoding, which can help them to structure the incoming information in a more abstract and explicitly accessible way. In our study, 7–10-year-old children were presented with a well-established probabilistic sequence learning task on two consecutive days, after receiving an instruction that on the second day, they would have to teach a peer about the task. During the task, we could simultaneously measure two different types of learning: model-free learning of local (lower-level) statistical correlations and model-based learning of the global (higher-level) statistical structures of the sequences. We predicted that in case the motivation to teach facilitates model-based encoding, children who received the instruction to teach would perform better in learning the higher-level statistical structures than children in the control group, who did not receive an instruction to teach. Furthermore, since previous studies showed competition between the two types of encoding processes during development, we also predicted that facilitating children's model-based learning will impair their model-free learning of the lower-level statistical correlations. Our results confirmed both predictions, showing an improved model-based higher-level structure learning and an impaired model-free lower-level statistical correlation learning in the Teaching Group, compared to the controls. Thus, prompting teaching affects children's encoding of the novel information, by biasing them to learn in a model-based way, which can help to build more abstract and explicitly accessible representations that could be shared with others.

1. Introduction

Teaching, defined as ‘an intentional activity that is pursued in order to increase the knowledge (or understanding) of another who lacks knowledge, has partial knowledge or possesses a false belief’ (Strauss, Ziv, & Stein, 2002, p.3.), is fundamental to the existence of cumulative culture, which is one of the main characteristics of our species. Without teaching, it would be impossible to transmit information and accumulate knowledge through generations (Strauss et al., 2002; Strauss,

Calero, & Sigman, 2014; Strauss & Ziv, 2012). Indeed, humans seem to be innately motivated to share their knowledge and teach new generations (Calero, Zylberberg, Ais, Semelman, & Sigman, 2015; Csibra & Gergely, 2009; Gergely & Csibra, 2006; Kruger & Tomasello, 1996; Strauss et al., 2014; Tomasello & Moll, 2010). Recent research brought extensive evidence that already from birth, humans are prepared to learn from each other, and that young infants are biased to acquire generic information in a fast and efficient way in the context of ostensive-communicative cues such as eye-contact, motherese or calling

[☆] This paper is a part of special issue “Special Issue in Honour of Jacques Mehler, Cognition’s founding editor”.

* Corresponding author.

E-mail address: hanna.marno@gmail.com (H. Marno).

¹ Senior authors.

<https://doi.org/10.1016/j.cognition.2021.104784>

Received 28 August 2020; Received in revised form 28 April 2021; Accepted 17 May 2021

0010-0277/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

their name (Csibra & Gergely, 2009; Gergely & Csibra, 2006). However, besides having the innate predisposition to learn from social partners via communication, humans also seem to have the natural cognitive ability to teach (Strauss & Ziv, 2012), and young children start to teach as a ‘natural instinct’ already before they would have a fully-fledged theory of mind (Calero, Goldin, & Sigman, 2018). For example, 12-month-old infants would spontaneously inform an adult about the location of an object the adult is looking for and was moved without the adult knowing (Liszkowski, Carpenter, Striano, & Tomasello, 2006; Liszkowski, Carpenter, & Tomasello, 2008), which can be considered as an example of ‘proto-teaching’ (Calero et al., 2018). Later, around the age of 3–4 while they already start to become more ostensive in pedagogically relevant moments, their early teaching mainly consists of demonstrations and non-verbal communication (Calero et al., 2015). Studies on children’s spontaneous teaching during their toddlerhood observed an increase in their use of ostensive cues and gestures, providing evidence that the capacity to teach emerges automatically and effortlessly during development (Calero et al., 2015). For example, at the age of 5 rather than using demonstrations, children’s dominant teaching strategy is to explicitly explain the rules of a game to an ignorant interlocutor (Bensalah, Olivier, & Stefaniak, 2012; Davis-Unger & Carlson, 2008a, 2008b; Strauss et al., 2002), and by the age of 6–8, they can already combine ostension, referential cues, and speech during teaching (Calero et al., 2018).

Thus, young children seem to be genuinely biased both to teach and to learn from others, which raises a further question: how can teaching potentially contribute to the development of later emerging cognitive skills during childhood? According to some studies, between the age of 3 and 6 years, children’s executive functioning is a significant predictor of their teaching efficacy, but at the same time, teaching seems to precede their theory of mind capacities (Davis-Unger & Carlson, 2008a). Indeed, numerous studies found a correlation between the development of teaching abilities and performance on theory of mind tasks in preschoolers (Bensalah et al., 2012; Davis-Unger & Carlson, 2008a, 2008b; Strauss et al., 2002). Thus, it might be that teaching precedes the development of many mental capacities, and in fact, teaching could be a driving force of later emerging cognitive skills (Csibra & Gergely, 2006, 2009, 2011). For example, during teaching, the teacher must have some self-awareness of the information s/he is about to convey, which means that the content of her/his knowledge needs to become accessible, therefore to some degree explicit (Calero et al., 2018).

But would this happen at the time of the retrieval of the information (during the act of teaching), or already at the time of the encoding (so that the information could be shared later)? While both of these possibilities may lead to some kind of representational changes in the content of the child’s knowledge, the second possibility would probably have stronger implications regarding the proposal that teaching might be a driving force of later emerging cognitive capacities of the child. If we assume that already young children are ready to share their knowledge, it is possible that the motivation to teach could already have an impact on how they encode novel information they encounter, i.e., by biasing them towards a more abstract encoding of the novel information, which can result in more explicit and accessible representations.

To test this possibility, we decided to implement a version of the Alternating Serial Reaction Time (ASRT) task (Howard Jr & Howard, 1997), which was designed to measure two types of learning simultaneously: model-free learning of local (lower-level) statistical regularities and model-based learning of global (higher-level) structures (Kóbor et al., 2018; Simor et al., 2019). While model-free learning refers to the encoding of raw occurrence probabilities that are experienced momentarily through the sensory input, during model-based learning the encoding of the information also relies on an internally stored structured model of the world that emerges based on past experience (Daw, Niv, & Dayan, 2005). By using such model-based encoding processes, it becomes possible to acquire highly complex knowledge of the world in an easier manner, compared to relying on solely model-free learning of the

sensory input (Nemeth, Janacsek, & Fiser, 2013). Thus, encoding new information through model-based learning is also more likely to result in more abstract and explicitly accessible representations of the information (Nemeth et al., 2013).

However, it is important to note that greater reliance on model-based learning processes that are related to internal models emerges only late in development, around the age of 12 (Janacsek, Fiser, & Nemeth, 2012). Moreover, according to some proposals the emerging functionality of model-based encoding around this age also signals a shift when learning processes adapt efficiently to more complex aspects of the world by relying more on internal model-based interpretations, while at the same time to some extent neglecting the raw probabilities of the sensory input, thus, relying less on model-free learning processes (Janacsek et al., 2012).

Indeed, a study of Nemeth et al. (2013) confirmed this proposal, by providing evidence that triggering model-based encoding processes in children at the age of 12 can lead to the impairment of their model-free learning of simple statistical properties of the sensory input. In their study, subjects of different age groups were instructed to complete the cued version of the ASRT task (Howard Jr & Howard, 1997) by responding to stimuli, which appeared according to a probabilistic sequence structure (e.g., 2r1r3r4r, where numbers represent four specific locations on the screen determined by the sequence, and r represent randomly selected location out of the four possible ones). Because of this probabilistic structure of the stimuli, two types of statistical regularities could be learned: local or lower-order statistical correlations and global or higher-level statistical structures. Furthermore, while encoding the lower-level statistical correlations could easily occur in a model-free learning manner, the acquisition of the global statistical structures become easier if the learning relies on model-based encoding processes, potentially resulting in also more explicitly accessible representations of the structure. In the study of Nemeth et al. (2013) subjects were either explicitly instructed to try to detect the global statistical structures of the sequences, or just performed the task without receiving such instruction. Results showed that the explicit instruction significantly helped all age groups in the detection of the higher-level statistical structures. Interestingly, however, while the explicit instruction to focus on the higher-level statistical structures also led to a slight improvement in the encoding of the lower-level statistical correlations in the older age groups, it had the opposite effect in the younger age groups, resulting in a lower performance in the encoding of the lower-level statistical correlations. Thus, when younger children were asked to focus on higher-level structures, this attentional switch impaired the way they encoded lower-level statistical correlations, potentially reflecting a competition between model-based and model-free encoding processes during development (Decker, Otto, Daw, & Hartley, 2016; Janacsek et al., 2012).

In the study of Nemeth et al. (2013) the crucial manipulation triggering a switch from model-free encoding processes to a model-based learning was the explicit instruction to try to detect the higher-level statistical structures of the stimuli sequences. However, if our hypothesis is correct, and the motivation to teach might facilitate children’s encoding of the incoming information in a more abstract and explicitly accessible manner, it is possible that - without giving them the explicit instruction to try to detect the global structures of the sequences - solely by instructing them to teach an ignorant interlocutor to perform the same task later might trigger a similar switch from model-free learning to a more model-based learning of the stimuli. Finding this effect, we would have the opportunity to directly witness how eliciting an early functioning motivation to teach can contribute to the development of children’s cognitive skills by biasing them towards using model-based encoding processes that only becomes fully-fledged later in life. Furthermore, from a methodological point of view, obtaining a similar dissociation in the learning performance that was found in the study of Nemeth et al. (2013) would also provide evidence that the instruction to teach does not only lead to a generally better performance due to

enhanced attention during the task, but it may potentially modify the encoding processes of young children.

Thus, in our study, we asked 7–10-year-old children to participate in the same cued version of the ASRT task that was used in the study of Nemeth et al. (2013). However, while one group of participants (i.e., the ‘Teaching Group’) was informed before the task that the next day they would have to teach everything they learned about the task to another person (who is known to be ignorant to the task), the other group did only receive the instruction to try to perform on the task as well as they could (the ‘Control Group’). Participants’ learning performance was measured twice: first during the training period, and then 24 h after the training. In addition, we also measured the explicit awareness of the sequences by asking them to explicitly report about the sequences after completing each block. We predicted that in case the motivation to teach would trigger a model-based encoding, resulting in potentially more explicitly accessible representations of the global structures of the task, participants in the Teaching Group would perform better in the detection of the higher-level statistical structures of the sequences, and also in the explicit awareness reports, than participants in the Control Group. However, since they are still in that stage of their development when their model-based encoding processes are not fully developed yet (Janacek et al., 2012; Nemeth et al., 2013), the improvement of the model-based encoding of the stimuli triggered by the teaching instruction might impair their model-free encoding of the lower-level statistical correlations. Therefore our second prediction was that subjects in the Teaching Group would perform worse in learning the lower-level statistical correlations compared to the Control Group. We aimed to test these hypotheses after a 24-h consolidation period as well because previous studies showed that group differences can be better detected after consolidation due to the stabilization of the acquired knowledge (Ambrus et al., 2020; Janacek, Ambrus, Paulus, Antal, & Nemeth, 2015; Vékony et al., 2020).

Finally, since we hypothesized that the effect would be there already at the time of encoding of the new information and not only at the time of the retrieval, we predicted that we would find a difference between the Teaching and the Control Group even if they would not perform any actual teaching (which would reflect an effect of the actual retrieval), but they would only need to keep in mind that later they would need to teach (which can affect the way of encoding of the information).²

2. Materials and methods

2.1. Participants

Sixty children between the age of 7 and 10 participated in the experiment ($M_{\text{age}} = 102.32 \pm 7.19$ months). Participants were randomly assigned to the Teaching or Control Group. The mean age of the participants in the two groups was equal ($M_{\text{teaching}} = 101.44 \pm 7.77$ months, $M_{\text{control}} = 103.18 \pm 6.63$ months). Five participants were excluded from the analysis because their mean accuracy or reaction times (RTs) were ± 2 standard deviations above or below the average of the full sample; thus, they were considered not to follow the instructions properly. The final sample contained data of 55 participants ($N_{\text{teaching}} = 27$, $N_{\text{control}} = 28$). None of the children suffered from any developmental, psychiatric, or neurological disorders. Parents gave informed consent and received no financial compensation for participation. The study was approved by the National Psychological Ethical Committee of Hungary (Ref. No.: 2017/61).

² Even though we did not assess participants’ teaching performance in our experiment, based on previous studies we know that children already after the age of 3 years can understand the instruction to teach, they usually positively react to it and also they manage to successfully teach a confederate (e.g. Davis-Unger & Carlson, 2008a, 2008b; Howe & Recchia, 2005; Strauss et al., 2002).

2.2. Tasks

2.2.1. Alternating Serial Reaction Time task (ASRT)

Learning was measured by the cued version of the Alternating Serial Reaction Time task (ASRT; Song, Howard, & Howard, 2007; Nemeth et al., 2013). In the task, a target stimulus (a drawing of a dog or a penguin) appeared in one of four horizontally arranged empty circles. From left to right, the Z, C, B, and M keys of a QWERTY keyboard corresponded to the four positions (all the remaining keycaps were removed). The participants had to press the key corresponding to the position of the target stimulus as fast and as accurately as they could (Fig. 1A). The target stimulus remained on the screen until the first correct keypress, and the next stimulus appeared after a 120 ms-long response-to-stimulus interval.

The task was divided into blocks of 85 stimuli: the first five stimuli appeared at a randomly selected position (for practice purposes), then an 8-element alternating sequence repeated ten times (e.g., 2r4r3r1r, where numbers represent the four circles on the screen from left to right and r represents a randomly selected circle). The regularity was marked by different stimuli for predefined (pattern) and random elements. In order to maintain the attention and motivation of the children, we used pictures of animals to indicate predefined/pattern (a dog’s head) and random elements (a penguin - for the random elements of the sequence and the first five stimuli). Participants were informed that appearance of the dog heads follow a predetermined pattern, while no information about the penguins were shared with them.

As a result of this alternating structure, some runs of three consecutive elements (henceforth referred to as *triplets*) occur with higher probability than others (Fig. 1B). For example, if the sequence was 2r1r3r4r, 2×1 , 4×2 , 3×4 , and 1×3 triplets („X” indicates the middle element of the triplet) occur with higher probability since their third elements could either be predefined or random. However, e.g., 2×3 and 4×1 triplets occur with lower probability since their third elements could only be random. The former triplet types are labeled as “high-probability triplets” (62.5% of all trials), while the latter types as “low-probability triplets” (37.5% of all trials).

Each trial can be categorized according to two aspects: 1) whether it was a pattern or random trial, and 2) whether it was the final element of a high- or low-probability triplet. Three different types of trials could be differentiated: pattern high-probability trials, random high-probability trials, and random low-probability trials (Fig. 1C). Because of the alternating sequence structure, all predefined (pattern) trials (50%) and by chance, one-fourth of random trials (12.5%) form high-probability triplets (62.5% in total), and the rest of the random trials form low-probability triplets (37.5%). For more details on the ASRT sequence structure, see „More details on the ASRT sequence-structure” subsection in the Supplementary Materials.

Based on the three different trial types, two types of learning can be differentiated. The difference in responses between random high-probability and random low-probability trials indicates the learning of *lower-level statistical correlations*. In this case, the stimuli’ sequence properties are controlled (because only random trials are analyzed). The only difference between the two stimulus types lies in the statistical probabilities (high- vs. low-probability). The difference in responses between pattern high-probability and random high-probability trials indicates the learning of *higher-level statistical structures*. Here, the statistical properties of the stimuli are controlled (both types of trials have a high probability of occurrence), and the only difference between the trials lies in their sequence properties (pattern vs. random trials).

2.2.2. Post-block explicit sequence report task

After each block of ASRT, the explicit knowledge about the sequence was measured. Participants were instructed to indicate the order of appearance of the dog heads (pattern stimuli) by pressing the corresponding response keys. This procedure lasted until participants gave 12 consecutive responses, then the ASRT task continued with the next

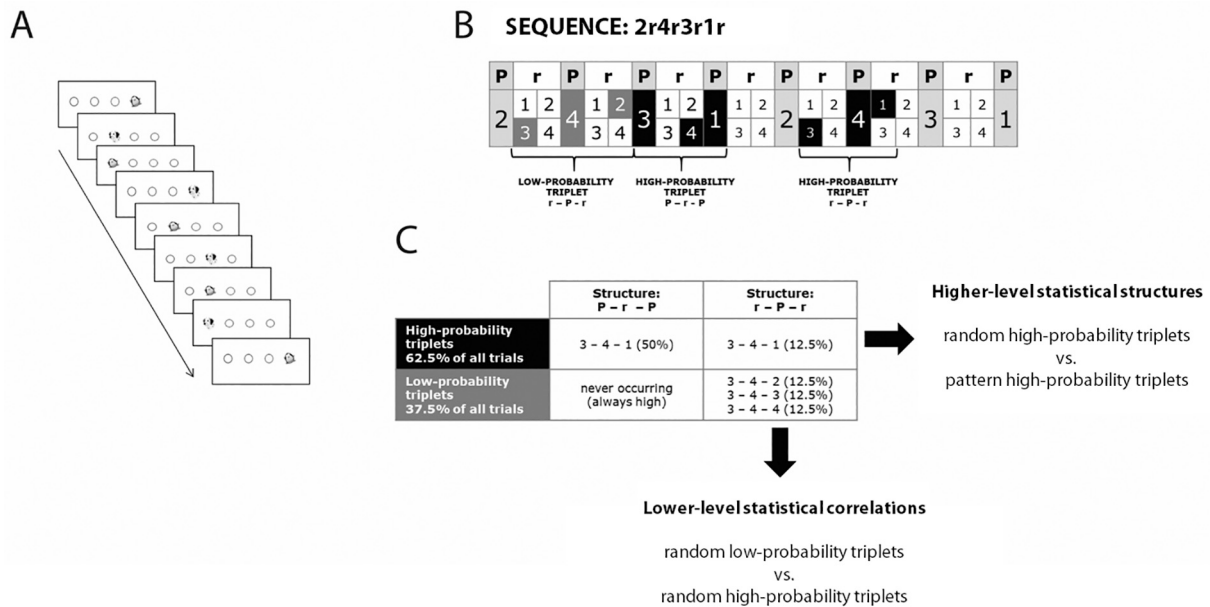


Fig. 1. The structure of the ASRT task. (A) Drawings of a dog's head or a penguin appeared on one of the four possible positions. The participants' task was to press the corresponding button as fast and as accurately as they could. (B) The ASRT task consisted of two types of stimuli: every first stimulus is a pattern element (P) and every second stimulus is a random (r) element. As a result of this structure, some triplets occur with higher probability than other triplets (high vs. low-probability triplets). (C) The high-probability triplets can be formed from two pattern elements and one random element in the middle (pattern high-probability triplets); however, high-probability triplets can be formed from two random and one pattern element (random high-probability triplets) too. Low-probability triplets can be formed from one pattern and two random elements. (Please note that each trial is categorized as one of the three trial types; that is, using a sliding window, each trial is defined as the third element of a pattern or random high-probability or a random low-probability triplet.) Higher-level statistical structure is defined by the difference between random high-probability and pattern high-probability triplets, and lower-level statistical correlations are defined by the difference between random high- and low-probability trials.

block. Ideally, participants reported the four repeating pattern elements of the sequence three times: if the sequence was 2r1r3r4r (where r indicates a random element), the entirely correct response was 213421342134. However, the sequence report was also considered entirely correct if only the starting point was different (for example, the sequences 2r1r3r4r and 1r3r4r2r result in the same triplet probabilities and are indistinguishable from each other).

2.3. Procedure

The experiment consisted of two sessions on two consecutive days. On the first day, participants completed 20 blocks of ASRT. Prior to learning, we told the members of the Teaching Group that they would play a game, and on the second day of the experiment, they will have to teach another child (who is ignorant to the task) everything they know about the game. The Control Group did not receive such instruction; they were asked to try to play the game as well as they could, and also that they would get later rewarded based on their performance. On the second day, participants completed five more blocks of ASRT with the same design. After completing the task on the second day, children in the Teaching Group were told that the 'ignorant' peer, in the end, could not come; therefore, there was no need to teach about the task.

2.4. Statistical analysis

2.4.1. Alternating Serial Reaction Time task

For the analysis of the ASRT task, we categorized each stimulus as either the third element of a high- or a low-probability triplet (depending on the positions of the $n-1$ and $n-2$ trials), and determined the reaction time (RT) of the response to this stimulus. As learning performance is mainly determined by RTs in ASRT studies, e.g., Janacsek, Borbély-Ipkovich, Nemeth, & Gonda, 2018; Virag et al., 2015; Howard Jr, Howard, Dennis, Yankovich, & Vaidya, 2004; Song et al.,

2007; Vékony et al., 2020), we report here the RTs only. However, all the analyses presented were performed on accuracy data, and results can be found in the Supplementary Materials. Only correct responses were considered for the analysis. Trills (e.g., 1-2-1, 4-3-4) and repetitions (e.g., 1-1-1, 4-4-4) were also eliminated as participants often show pre-existing response tendencies to them (Howard Jr et al., 2004).

Five consecutive blocks were analyzed instead of single blocks to facilitate data analysis and improve statistical power. The units of analysis are referred to as Day 1 B1-B5, Day 1 B6-B10, Day 1 B11-B15, Day 1 B16-B20, indicating the blocks on the first day, and Day 2 B1-B5 indicates the blocks on the second day. We calculated median RTs for correct responses for each participant and each unit of five blocks, separately for pattern high-probability, random high-probability, and random low-probability trials. The median RTs were submitted to mixed-design ANOVAs to evaluate learning and the consolidation of knowledge for the two types of learning (lower-order vs. higher-order statistical structure). 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_p^2) as the measure of effect size. Bonferroni correction was used for pairwise comparisons.

2.4.2. Post-block explicit sequence report task

To evaluate the participants' performance on the post-block explicit sequence report task, we computed two indices. For the first index, we calculated the accuracy of the sequence report after each block (*percentage score of explicit knowledge*). To this aim, we scored each keypress as a correct or incorrect keypress relative to the position of the first keypress to check if the participants could accurately report the sequence. We then calculated the percentage of the correct keypresses after each block. Here, the possible values ranged from 0% to 100% after each block (0% meant that no keypress was correct after the first keypress; 100% performance meant that all keypresses were correct after the first keypress, that is, the series of keypresses accurately reflected the

sequence embedded in the ASRT task). The average of the 20 percentage scores (one percentage score after each block, and we had 20 blocks) was calculated for each participant. Henceforth, this index is referred to as the *percentage score of explicit knowledge*. Here, the higher explicit knowledge percentage score indicates more stable explicit knowledge about the sequence structure.

For the second index – *number of accurate responders* –, after each block, we assessed whether the participant reported the sequence with 100% accuracy (all keypresses were accurate). We then counted the number of blocks in which the given participant reported the sequence with 100% accuracy. Here, for each participant, the possible scores could have ranged from 0 (meaning the participant did not respond with 100% accuracy in any of the blocks) to 20 (meaning that the participants responded with 100% accuracy in all blocks). For each group, we counted the number of participants who performed the sequence report task with 100% accuracy at least ten out of the 20 blocks. Henceforth, this index is referred to as the *number of accurate responders*. Here, the higher number of participants in one of the two groups indicates more stable group-level explicit knowledge.

The explicit knowledge of the sequence was evaluated with an independent samples *t*-test to compare the percentage score of explicit knowledge between the two groups. The number of accurate responders in the two groups was compared with a chi-squared test.

3. Results

3.1. Learning of lower-level statistical correlations on Day 1

To compare the learning of lower-level statistical correlations of the two groups in the first session, we performed a mixed-design ANOVA with the within-subject factor of Blocks (Day 1 B1-B5 vs. B6-B10 vs. B11-B15 vs. B16-B20) and Triplet (random high-probability vs. random low-probability trials) and the between-subject factor of Group (Teaching Group vs. Control Group). The ANOVA revealed a significant main effect of Block, $F_{3,159} = 75.62, p < .001, \eta_p^2 = 0.59$, indicating that the RTs on

average became smaller over the course of learning. The main effect of Group was not significant, $F_{1,53} = 0.23, p = .64, \eta_p^2 = 0.004$, indicating that, irrespective of stimulus type, average RTs of the two groups were similar. The interaction between Block and Group was not significant either, $F_{3,159} = 0.35, p = .75, \eta_p^2 = 0.007$. It indicates that there was no statistically significant difference between groups in how the average RTs changed throughout the task (that is, no difference in their general skill learning).

The main effect of Triplet was significant, $F_{1,53} = 112.145, p < .001, \eta_p^2 = 0.68$, revealing that participants were faster on random high-probability trials ($M = 597.52 \text{ ms} \pm 12.12 \text{ ms}$) than on random low-probability trials ($M = 627.77 \text{ ms} \pm 12.42 \text{ ms}$); thus, learning of lower-level statistical correlations occurred. The interaction of Block and Triplet factors was also not significant, $F_{3,159} = 0.96, p = .41, \eta_p^2 = 0.02$, indicating that the degree of learning of lower-level statistical correlations did not change throughout the task. More importantly, the interaction of Group and Triplet was also not significant, $F_{1,53} = 0.79, p = .38, \eta_p^2 = 0.02$, indicating that the two groups did not differ in terms of their knowledge of lower-level statistical correlations in the first session. The non-significant interaction of the Group, Triplet, and Block factors reveals that this pattern did not change over the course of the first session, $F_{3,159} = 0.31, p = .82, \eta_p^2 = 0.006$ (Fig. 2).

3.2. Learning of higher-level statistical structures on Day 1

To compare the performance in terms of learning of higher-level statistical structures, we performed a mixed-design ANOVA with the within-subject factor of Blocks of Day 1 (B1-B5 vs. B6-B10 vs. B11-B15 vs. B16-B20) and Triplet (pattern high-probability vs. random high-probability trials), and the between-subject factor of Group (Teaching Group vs. Control Group). We found a significant main effect of Block, $F_{3,159} = 90.09, p < .001, \eta_p^2 = 0.63$, indicating again that the RTs on average became smaller over the course of learning. The main effect of Group was not significant, $F_{1,53} = 0.092, p = .76, \eta_p^2 = 0.002$, indicating that, irrespective of stimulus type, the average RTs of the two groups

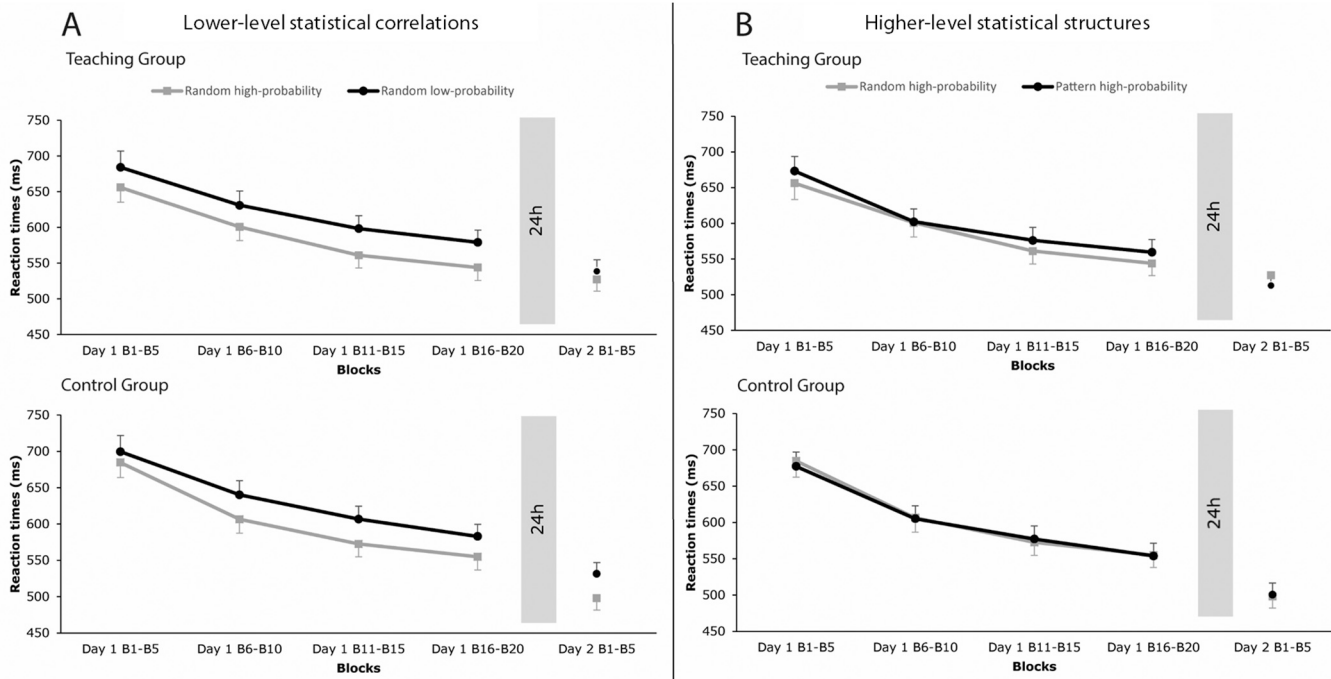


Fig. 2. Learning of lower-level statistical correlations (A) and learning of higher-level statistical structures (B) in the Teaching and Control Group. The vertical axis indicates the reaction times in milliseconds, and the horizontal axis the blocks of ASRT in Day 1 and Day 2. In learning of lower-level statistical correlations, no difference was found between groups. In learning of higher-level statistical structures, the Teaching Group showed marginally significant advantage in differentiating between random and pattern high-probability triplets compared to the Control Group. The error bars represent the standard error of the mean (SEM).

were similar. The interaction between Block and Group was not significant either, $F_{3,159} = 0.30$, $p = .77$, $\eta_p^2 = 0.006$, suggesting the lack of statistically significant difference between the groups in how the average RTs changed throughout the task.

The main effect of Triplet was not significant, $F_{1,53} = 2.55$, $p = .116$, $\eta_p^2 = 0.05$, revealing that, overall, participants did not exhibit learning of higher-level statistical structures in the first session. The interaction of Block and Triplet was not significant either, $F_{3,159} = 0.51$, $p = .68$, $\eta_p^2 = 0.01$, indicating that the degree of the (lack) of learning of higher-level statistical structures did not change over the course of the task. However, the interaction of Group and Triplet was marginally significant, $F_{1,53} = 3.85$, $p = .055$, $\eta_p^2 = 0.07$ (difference $M_{Teaching} = 12.26 \pm 4.91$ vs. $M_{Control} = 1.25 \pm 4.82$). The non-significant interaction of the Triplet, Block, and Group factors showed that degree of knowledge of higher-level statistical structures did not change over time between the two groups, $F_{3,159} = 0.63$, $p = .59$, $\eta_p^2 = 0.01$ (Fig. 2).

3.3. Changes in knowledge of lower-level statistical correlations from Day 1 to Day 2

To reveal how knowledge of lower-level statistical correlations consolidated in the two groups, we ran a mixed-design ANOVA with the within-subject factor of Block (Day 1 B16-B20 vs. Day 2 B1-B5) and Triplet (random high-probability vs. random low-probability triplets), and the between-subject factor of Group (Teaching Group vs. Control Group). Please note that the comparisons including the Triplet factor, reveals changes in knowledge of lower-level statistical correlations. In contrast, the comparisons excluding the Triplet factor indicate changes in the offline general skill improvements. The main effect of Block was significant, $F_{1,53} = 38.24$, $p < .001$, $\eta_p^2 = 0.42$, indicating that, irrespective of the stimulus type, participants became faster between sessions. The interaction of Block and Group was marginally significant, $F_{1,53} = 3.64$, $p = .062$, $\eta_p^2 = 0.06$, revealing that, on trend-level, the Control Group became faster from Day 1 to Day 2 ($M_{difference} = -18.27$ ms \pm 22.43 ms) to a greater extent than the Teaching Group ($M_{difference} = -7.31$ ms \pm 24.05 ms).

The main effect of Triplet was significant, $F_{1,53} = 44.13$, $p < .001$, $\eta_p^2 = 0.45$: participants were faster on random high-probability trials compared to random low-probability trials, indicating knowledge of lower-level statistical correlations. The interaction of Triplet and Block was not significant, $F_{1,53} = 1.77$, $p = .19$, $\eta_p^2 = 0.03$, indicating that, on average, learning of lower-level statistical correlations did not change over the offline period. The interaction of Triplet and Group was not significant, $F_{1,53} = 0.82$, $p = .37$, $\eta_p^2 = 0.02$, indicating that, overall, knowledge of lower-level statistical correlations did not differ between the two groups. More importantly, the three-way interaction of Triplet, Block, and Group was significant, $F_{1,53} = 4.31$, $p = .04$, $\eta_p^2 = 0.08$, indicating that the degree of knowledge of lower-level statistical correlations changed differently over the offline period in the two groups. The pairwise comparisons revealed that, in the Teaching Group, knowledge of lower-level statistical correlations was detectable in Day 1 B15-B20 (random high vs. random low-probability trials: $M = 35.26$ ms \pm 8.96, $p < .001$), and it became smaller in Day 2 B1-B5 ($M = 11.46$ ms \pm 6.07, $p = .065$). However, in the Control Group, knowledge of lower-level statistical correlations was revealed at both time-points (Day 1: $M = 28.13$ ms \pm 8.80, $p = .002$; Day 2: $M = 33.32$ ms \pm 5.96, $p < .001$).

As a follow-up of the significant three-way interaction, we subtracted the RTs for random high-probability trials from the RTs for random low-probability trials. We repeated the analysis with this learning score as the dependent variable. The ANOVA revealed that the learning of lower-level statistical correlations score significantly changed from Day 1 to Day 2 in the Teaching Group ($p = .02$), but not in the Control Group ($p = .60$) (Fig. 3).

3.4. Changes in knowledge of higher-level statistical structures from Day 1 to Day 2

A mixed-design ANOVA was performed to reveal how knowledge of higher-level statistical structures consolidated in the two groups: a 2 (Block: Day 1 B16-B20 vs. Day 2 B1-B5) \times 2 (Triplet: pattern high-probability vs. random high-probability trials) \times 2 (Group: Teaching Group vs. Control Group) ANOVA was performed. Please note that contrary to the analysis for lower-level statistical correlations, the Triplet factor contained the pattern of high-probability vs. random high-probability triplets. Again, the comparisons including the Triplet factors reveal changes in learning the higher-level statistical structures, while the comparisons excluding the Triplet factors indicate changes in the offline general skill improvements. The main effect of Block was significant, $F_{1,53} = 50.75$, $p < .001$, $\eta_p^2 = 0.49$, indicating that irrespective of the stimulus type, participants were faster on Day 2 than on Day 1. The interaction of Block and Group was marginally significant, $F_{1,53} = 3.68$, $p = .060$, $\eta_p^2 = 0.07$, indicating that, on trend-level, the Control Group became faster from Day 1 to Day 2 ($M_{difference} = -55.13$ ms \pm 8.55 ms) to a greater extent than the Teaching Group ($M_{difference} = -31.74$ ms \pm 8.70 ms).

The main effect of Triplet was not significant, $F_{1,53} = 0.22$, $p = .88$, $\eta_p^2 < 0.01$: thus, overall, no evidence was found for learning of higher-level statistical structures. The interaction of Triplet and Block was significant, $F_{1,53} = 5.75$, $p = .02$, $\eta_p^2 = 0.10$: on average, participants improved more for pattern than for random high-probability trials over the two days. The interaction of Triplet and Group was not significant, $F_{1,53} = 0.001$, $p = .98$, $\eta_p^2 < 0.02$, indicating that, overall, knowledge of higher-level statistical structures did not differ between the two groups. Importantly, the three-way interaction of Triplet, Block, and Group was also significant, $F_{1,53} = 9.08$, $p = .004$, $\eta_p^2 = 0.15$, indicating that the degree of knowledge of higher-level statistical structures changed differently in the two groups over offline period. The pairwise comparisons revealed that in the Teaching Group, knowledge of higher-level statistical structures was detectable in Day 2 B1-B5 (random high- vs. pattern high-probability trials: $M = 14.43$ ms \pm 6.88, $p = .04$) but not in the Day 1 B15-B20 ($M = -15.54$ ms \pm 8.12, $p = .06$). However, in the Control Group, knowledge of higher-level statistical structures was detected at neither time-points (Day 1: $M = 0.93$ ms \pm 7.98, $p = .91$; Day 2: $M = -2.48$ ms \pm 6.76, $p = .72$) (Fig. 2).

As a follow-up of the significant three-way interaction, we subtracted the RTs for pattern high-probability trials from the RTs for random high-probability trials. We repeated the analysis with this learning score as the dependent variable. The ANOVA revealed that the learning scores changed significantly from Day 1 to Day 2 in the Teaching Group ($p < .01$) but not in the Control Group ($p = .66$) (Fig. 3).

3.5. Reports of the explicit knowledge of the sequence

We compared the number of accurate responders between the two groups. The chi-square test revealed that in the Teaching Group, more children (16 out of 27) were able to report the sequence with 100% accuracy in at least half of the blocks than in the Control Group (9 out of 28), $\chi^2(1, N = 55) = 4.08$, $p = .04$.

Furthermore, to compare the awareness of the sequence structure between the two groups, we performed an independent samples *t*-test on the percentage score of explicit knowledge of the post-block sequence report task (for details, see the 'Sequence report task' subsection of the Methods section). The *t*-test revealed a trend-level advantage for the Teaching Group, $t(53) = 1.728$, $p = .09$ ($M_{Teaching} = 68.36\% \pm 28.52\%$, $M_{Control} = 55.37\% \pm 27.23\%$).

4. Discussion

By using a modified version of a well-established probabilistic sequence learning task, we tested the hypothesis that instructing young

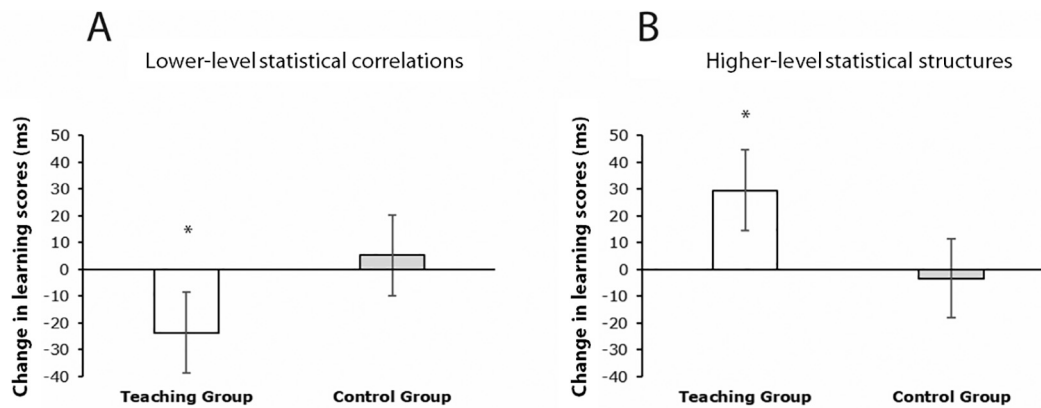


Fig. 3. The change in learning scores from Day 1 to Day 2 in learning of lower-level statistical correlations (A) and learning of higher-level statistical structure (B). The vertical axes indicate the change in learning scores in milliseconds (for learning of lower-level statistical correlations: random low-probability minus random high-probability trials on Day 2 minus those on Day 1; for learning of higher-level statistical correlations: random high-probability trials minus pattern high-probability trials on Day 2 minus those on Day 1). Thus, bars greater than zero could indicate an increase in knowledge over the offline period, and bars smaller than zero could indicate a decrease in knowledge. The horizontal axes indicate the two groups. A decrease in learning scores for lower-level statistical correlations was observed in the Teaching Group with no change in the Control Group. In contrast, for learning of higher-level statistical structures, an increase in learning scores was observed in the Teaching Group, with no changes in the Control Group. The error bars represent SEM. * $p < .05$

children to teach an ignorant interlocutor can have an impact on the way they encode novel information. Specifically, we predicted that prompting teaching will facilitate model-based learning processes, while potentially hinder model-free encoding processes. To this end, we measured their learning performance for two types of information: local or lower-level statistical correlations and global or higher-level statistical structures of stimuli sequences.

On the first day in both groups children learned the lower-level statistical correlations of the sequences to the same extent. However, compared to the first day, learning performance for the lower-level statistical correlations became smaller on the second day in the Teaching Group, while the Control Group showed similar level of knowledge between the two sessions. Regarding the learning of the higher-level statistical structures of the sequences, no significant difference emerged between the two groups on the first day. However, on the second day, the Teaching Group but not the Control Group revealed significant learning performance for the higher-level statistical structures of the sequences. Finally, when we asked children to report about the global structures of the sequence explicitly, we found that children in the Teaching Group showed better performance in the post-block sequence report task than the Control Group. In sum, while the Control Group did not show any intentional learning of the higher-order statistical structure information, the Teaching Group revealed significant improvement after a consolidation period and also managed to report better about the structures explicitly. Regarding the learning performance for the lower-level statistical correlations, however, the Teaching Group exhibited some *forgetting* by the second day, while The Control Group's learning performance was continuous on both days.

Although children were told that some stimuli follow a pre-determined order, they were not explicitly instructed to use that information to improve their performance, in contrast to the study of Nemeth et al. (2013). Thus, even though children in our study were not explicitly told to focus on the higher-level statistical structures of the sequences, the instruction to teach an ignorant interlocutor about the task resulted

in a similar dissociation in their encoding processes that was observed in the Nemeth et al. study (Nemeth et al., 2013), leading to an enhanced model-based encoding of the higher-level statistical structures, at the expense of their model-free encoding of the lower-level statistical correlations. Furthermore, it is important to note that this dissociation occurred even though children never actually performed any teaching; that is, they only knew that on the second day, they would have to teach.³ Hence, the effect could not be due to the retrieval of the information, but it was already present during the time of encoding, and it became more expressed after a consolidation period (on the second day).

The idea that teaching could be a driving force of later emerging cognitive capacities of the child has already been raised in the past (Bensalah et al., 2012; Calero et al., 2018; Csibra & Gergely, 2006, 2009, 2011; Strauss et al., 2002), but studies so far mainly focused on correlational relations between children's teaching capacities and other cognitive skills, such as theory of mind or executive functions (Bensalah et al., 2012; Davis-Unger & Carlson, 2008a, 2008b; Strauss et al., 2002). Our study provided evidence that by triggering the motivation to teach, children start to encode novel information in a more abstract, model-based manner, which can lead to explicitly accessible representation of the information. Furthermore, even though this type of learning is usually not fully developed yet before the age of 12 years (Janacsek et al., 2012; Nemeth et al., 2013), the instruction to teach could facilitate model-based learning already in children between the age of 7 and 10 years. Thus, we can conclude that the motivation to teach seems to have a direct impact on the development of the later emerging cognitive capacity to encode novel information in a more abstract, model-based manner. While it is a further question whether this type of encoding may contribute to the development of other cognitive capacities as well (for example theory of mind or metacognition, which are usually considered as necessary skills for fully-fledged teaching abilities), to our knowledge our study provides the first piece of evidence that children's motivation to teach can have a direct impact on their cognitive development.

³ Since the primary aim of our study was to test whether children in the experimental group, simply by being aware that later they would need to teach would encode the sequence information differently from the control group, we did not assess their actual teaching performance. However, an interesting question for future studies could be whether children, who manage to encode the global structure information better would also be more successful at teaching this information to their peers later.

One could also speculate that the effect we found was not specifically due to the instruction to teach, but due to the mere fact that by instructing children to teach, they were prompted to encode the perceived information in a way that they could verbalize it, which led to a selective encoding of the global or higher-level structure information (which could be verbalized at least partially, in contrast to the lower-level statistical correlations). While we cannot exclude this possibility, we would like to point out that since the act of verbalization is a necessary and preferred mean to share information with others (and indeed, by the age of 5 children start to switch from providing demonstrations to verbal explanations; [Strauss & Ziv, 2012](#)), the underlying motivation of verbalization is usually the same social motivation of the one of teaching: to share our knowledge with others. From this point of view, even if we assume that the perceived dissociation in the encoding of the novel information was directly related to the aim to be able to verbalize the information, we propose that the capacity to verbalize could originate from the motivation to be able to share the information with the other, which is also the underlying motivation of teaching.

Studying the impact of teaching on the knowledge of the teacher has been a somewhat neglected field in the literature, but recently numerous studies have been conducted on a related question: the effect of explanation on children's causal learning and abstract reasoning (e.g., [Legare & Lombrozo, 2014](#); [Legare, Sobel, & Callanan, 2017](#); [Walker, Lombrozo, Legare, & Gopnik, 2014](#); [Williams & Lombrozo, 2010](#)). In these studies, children are typically asked to provide explanations during solving certain tasks (for example, while they observe how different objects can activate a toy, they generate explanations regarding the causal affordances of the objects). According to these experiments, the act of explanation leads children to a better understanding of the underlying causalities of the observed mechanism and also to a more efficient generalization of the causal role they learned. As the authors argue, explaining their observations seems to encourage children to focus on causal mechanisms and generalization, while it also helps them to avoid superficial, perceptually-bound judgments during making decisions about category membership ([Walker et al., 2014](#)). This is because, invoking mechanisms and broad generalizations are characteristics of good explanations ([Legare & Lombrozo, 2014](#)), and explanation can drive the discovery of regularities and generalize it to novel contexts, sometimes even resulting in overgeneralizations by hindering exceptions ([Walker, Williams, Lombrozo, & Gopnik, 2012](#); [Williams & Lombrozo, 2010, 2013](#); [Williams, Lombrozo, & Rehder, 2013](#)). Thus, as the authors conclude, explaining may be intimately related to learning new concepts and theories ([Williams & Lombrozo, 2010](#)).

However, since the explanation is the common way of verbally conveying knowledge, it is possible that the facilitatory effect of explanation on children's causal learning processes is also related to their social motivation to share their knowledge with others, that is, to teach. Therefore it might be that beyond helping the discovery of causalities and promoting generalizations, explanation can contribute in general to the development of later emerging cognitive capacities of the child by facilitating the encoding of information in explicit, accessible representations and by promoting self-awareness regarding the knowledge of the child. Indeed, children's explanations seem to contribute not only to their causal reasoning about the physical world but also to their understanding of biological and social-conventional phenomena in general ([Hickling & Wellman, 2001](#)). Furthermore, the facilitatory effect of explanation on children's causal reasoning seems to disappear when instead of providing explanations, children are simply asked to 'think aloud' about the observed events ([Williams & Lombrozo, 2010](#)). Thus, when the act of verbalizing lacks the social motivation to share the child's knowledge with the other, children do not exhibit any improvement in their learning processes. This suggests that beyond merely verbally describing the novel information, children also need to address someone with the intention to convey their knowledge in order to achieve further representational changes.

Since our study aimed to investigate a somewhat less examined

question of the field, how teaching might have an impact on the knowledge of the 'teacher' and contribute to the development of further cognitive capacities, we are aware of the fact that our results should be interpreted in a cautious way. However, we hope that as a first step, we managed to provide evidence that prompting teaching does have an impact on children's encoding of the novel information, in this case by triggering a model-based encoding that might result in more explicit, accessible representations that could be shared with others. A next step should be to investigate what are exactly the cognitive processes that contributed to a modified encoding of the information due to the instruction to teach, and also the implementation of further paradigms in order to validate the effect we found.

In sum, we believe that these findings could be a good start for further investigations on how this change of information encoding and representation can contribute to the development of other cognitive capacities during childhood.

Data availability

Data available at https://osf.io/5ky3u/?view_only=d65c3661e26f433d8c8e5341f73329e6.

Author contributions

Hanna Marno: Conceptualization, Methodology, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision. **Róbert Danyi:** Formal analysis, Investigation, Writing - Original Draft. **Teodóra Vékony:** Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization. **Karolina Janacsek:** Software, Methodology, Writing - Review & Editing, Supervision. **Dezso Németh:** Methodology, Formal analysis, Writing - Review & Editing, Supervision.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgment

We thank to J. Fiser and Y. Vidal for their useful comments on the manuscript. This research was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant 609819 (Somics project); the National Brain Research Program (project 2017-1.2.1-NKP-2017-00002); Hungarian Scientific Research Fund (NKFIH-OTKA K 128016, PI: D. N., NKFIH-OTKA PD 124148, PI: K.J.); Janos Bolyai Research Fellowship of the Hungarian Academy of Sciences (to K. J.); IDEXLYON Fellowship of the University of Lyon as part of the Programme Investissements d'Avenir (ANR-16-IDEX-0005) (to D.N.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2021.104784>.

References

- Ambrus, G. G., Vékony, T., Janacsek, K., Trimborn, A. B., Kovács, G., & Németh, D. (2020). When less is more: Enhanced statistical learning of non-adjacent dependencies after disruption of bilateral DLPFC. *Journal of Memory and Language*, *114*, 104144. <https://doi.org/10.1016/j.jml.2020.104144>.
- Bensalah, L., Olivier, M., & Stefaniak, N. (2012). Acquisition of the concept of teaching and its relationship with theory of mind in French 3- to 6-year-olds. *Teaching and Teacher Education*, *28*, 303–311. <https://doi.org/10.1016/j.tate.2011.10.008>.
- Calero, C. I., Goldin, A. P., & Sigman, M. (2018). The teaching instinct. *Review of Philosophy and Psychology*, *9*(4), 819–830. <https://doi.org/10.1007/s13164-018-0383-6>.

- Calero, C. I., Zylberberg, A., Ais, J., Semelman, M., & Sigman, M. (2015). Young children are natural pedagogues. *Cognitive Development*, 35, 65–78. <https://doi.org/10.1016/j.cogdev.2015.03.001>.
- Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *Processes of change in brain and cognitive development. Attention and performance XXI*, 21, 249–274.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153. <https://doi.org/10.1016/j.tics.2009.01.005>.
- Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 366(1567), 1149–1157. <https://doi.org/10.1098/rstb.2010.0319>.
- Davis-Unger, A. C., & Carlson, S. M. (2008a). Development of teaching skills and relations to theory of mind in preschoolers. *Journal of Cognition and Development*, 9(1), 26–45. <https://doi.org/10.1080/15248370701836584>.
- Davis-Unger, A. C., & Carlson, S. M. (2008b). Children's teaching: Relations to theory of mind and executive function. *Mind, Brain, and Education*, 2, 128–135. <https://doi.org/10.1111/j.1751-228X.2008.00043.x>.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8, 1704–1711. <https://doi.org/10.1038/nn1560>.
- Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From creatures of habit to goal-directed learners: Tracking the developmental emergence of model-based reinforcement learning. *Psychological Science*, 27(6), 848–858.
- Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. *Roots of human sociality: Culture, cognition, and human interaction*, 229–255.
- Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology*, 37(5), 668–683. <https://doi.org/10.1037/0012-1649.37.5.668>.
- 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. <https://doi.org/10.1037/0882-7974.12.4.634>.
- Howard, J. H., Jr., Howard, D. V., Dennis, N. A., Yankovich, H., & Vaidya, C. J. (2004). Implicit spatial contextual learning in healthy aging. *Neuropsychology*, 18(1), 124.
- Howe, N., & Recchia, H. (2005). Playmates and teachers: Reciprocal and complementary interactions between siblings. *Journal of Family Psychology*, 19(4), 497. <https://doi.org/10.1037/0893-3200.19.4.497>.
- Janacek, K., Ambrus, G. G., Paulus, W., Antal, A., & Nemeth, D. (2015). Right hemisphere advantage in statistical learning: Evidence from a probabilistic sequence learning task. *Brain Stimulation*, 8(2), 277–282. <https://doi.org/10.1016/j.brs.2014.11.008>.
- Janacek, 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. <https://doi.org/10.1016/j.pnpbp.2017.09.021>.
- Janacek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, 15(4), 496–505. <https://doi.org/10.1111/j.1467-7687.2012.01150.x>.
- Kóbor, A., Takács, Á., Kardos, Z., Janacek, 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. <https://doi.org/10.1016/j.biopsycho.2018.04.001>.
- Kruger, A. C., & Tomasello, M. (1996). Cultural learning and learning culture. In *The handbook of education and human development* (pp. 369–387).
- Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology*, 126, 198–212. <https://doi.org/10.1016/j.jecp.2014.03.001>.
- Legare, C. H., Sobel, D. M., & Callanan, M. (2017). Causal learning is collaborative: Examining explanation and exploration in social contexts. *Psychonomic Bulletin & Review*, 24(5), 1548–1554. <https://doi.org/10.3758/s13423-017-1351-3>.
- Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, 7(2), 173–187. https://doi.org/10.1207/s15327647jcd0702_2.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, 108(3), 732–739. <https://doi.org/10.1016/j.cognition.2008.06.013>.
- Nemeth, D., Janacek, K., & Fiser, J. (2013). Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Frontiers in Computational Neuroscience*, 7, 147. <https://doi.org/10.3389/fncom.2013.00147>.
- Simor, P., Zavecz, Z., Horváth, K., Élterő, 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. <https://doi.org/10.3389/fpsyg.2018.02708>.
- Song, S., Howard, J. H. J., & Howard, D. V. (2007). Implicit probabilistic sequence learning is independent of explicit awareness. *Learning & Memory*, 14, 167–176. <https://doi.org/10.1101/lm.437407>.
- Strauss, S., Calero, C. I., & Sigman, M. (2014). Teaching, naturally. *Trends in Neuroscience and Education*, 3(2), 38–43. <https://doi.org/10.1016/j.tine.2014.05.001>.
- Strauss, S., & Ziv, M. (2012). Teaching is a natural cognitive ability for humans. *Mind, Brain, and Education*, 6(4), 186–196. <https://doi.org/10.1111/j.1751-228X.2012.01156.x>.
- Strauss, S., Ziv, M., & Stein, A. (2002). Teaching as a natural cognition and its relations to preschoolers' developing theory of mind. *Cognitive Development*, 17(3), 1473–1487. [https://doi.org/10.1016/S0885-2014\(02\)00128-4](https://doi.org/10.1016/S0885-2014(02)00128-4).
- Tomasello, M., & Moll, H. (2010). The gap is social: Human shared intentionality and culture. In *Mind the gap* (pp. 331–349). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-02725-3_16.
- Vékony, T., Marossy, H., Must, A., Vécsei, L., Janacek, K., & Nemeth, D. (2020). Speed or accuracy instructions during skill learning do not affect the acquired knowledge. *Cerebral Cortex Communications*, 1(1). <https://doi.org/10.1093/texcom/tgaa041>.
- Vékony, T., Török, L., Pedraza, F., Schipper, K., Pleche, C., Tóth, L., ... Nemeth, D. (2020). Retrieval of a well-established skill is resistant to distraction: Evidence from an implicit probabilistic sequence learning task. *PLoS One*, 15(12). <https://doi.org/10.1371/journal.pone.0243541>.
- Virag, M., Janacek, K., Horvath, A., Bujdosó, Z., Fabo, D., & Nemeth, D. (2015). Competition between frontal lobe functions and implicit sequence learning: Evidence from the long-term effects of alcohol. *Experimental Brain Research*, 233(7), 2081–2089. <https://doi.org/10.1007/s00221-015-4279-8>.
- Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, 133(2), 343–357. <https://doi.org/10.1016/j.cognition.2014.07.008>.
- Walker, C. M., Williams, J. J., Lombrozo, T., & Gopnik, A. (2012). Explaining influences children's reliance on evidence and prior knowledge in causal induction. In *Proceedings of the annual meeting of the cognitive science society* (vol. 34, no. 34).
- Williams, J. J., & Lombrozo, T. (2010). Explanation constrains learning, and prior knowledge constrains explanation. In *Proceedings of the annual meeting of the cognitive science society* (vol. 32, no. 32).
- Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, 66(1), 55–84. <https://doi.org/10.1016/j.cogpsych.2012.09.002>.
- Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*, 142(4), 1006. <https://doi.org/10.1037/a0030996>.