Non-linear effects of presentation rate on sequence learning

Schmid, Samuel; Saddy, Doug; Franck, Julie

Non-linear effects of presentation rate on sequence learning

Samuel Schmid,[a] Douglas Saddy,[b] Julie Franck[a]

[a]*Faculty of Psychology, University of Geneva, Geneva, Switzerland*

[b]*Centre for Integrative Neuroscience and Neurodynamics, University of Reading, Reading, United Kingdom*

Abstract

In this article, we explore the impact of presentation rate on the extraction of hierarchical structure by manipulating the duration of the Response-to-Stimulus Interval (RSI) in a Serial Reaction Time (SRT) task. Multiple hypotheses have been put forward in the literature to account for the influence of RSI duration on sequence learning in the SRT task (Frensch & Miner, 1994; Huang et al., 2017; Willingham et al., 1997). However, this question has never been addressed from the perspective of hierarchical structure extraction. We found that RSI duration affected hierarchical elaboration in a non-linear way, with participants building higher hierarchical structures with an RSI of 250 ms compared to RSIs of 1000 ms and 100 ms. This finding suggests the presence of an optimal temporal window for sequence learning in the SRT task. This U-shaped effect cannot be accounted for by any of the existing hypotheses on the influence of RSI duration on sequence learning in the SRT task. We hypothesized that this effect results from the tension between the cognitive system's limited encoding capacity and the amount of information per unit of time delivered to the system.

Introduction

The understanding of the mechanism allowing the acquisition of the regularity underlying a sequence of events represents a fundamental question in cognitive science. Among the multiple methods adopted to explore sequence learning, the Serial Reaction Time (SRT) task is one of the most prominent. In this task, the elements of the sequence are presented one by one and the sequence determines the position of a stimulus on a screen. On each trial, the participant must press the button associated with the position of the stimulus as quickly as possible. Once the response is made, the stimulus disappears and the next trial begins. In the initial version of Nissen and Bullemer (1987), the stimulus could appear in four possible positions; this setting has been widely adopted in following studies. The assessment of sequence learning is done in two phases. In the learning phase, participants perform several blocks where the order of appearance of the stimuli is determined by the target sequence to be learned. This phase is followed by a so-called *transfer block* where the order of the stimuli is determined by an alternative sequence and then again by a block following the target sequence. The target sequence is considered as learned if the participants show a *transfer effect*, i.e., a slowing down in the transfer block compared to the adjacent sequenced blocks (Schwarb & Schumacher, 2012). The target sequence is typically 10-12 items long and is presented in a loop through the blocks. The most commonly used type of sequence follows a Second Order Conditional (SOC) structure (Reed & Johnson, 1994). In SOC sequences, the position of a trial t_n can be predicted with certainty by the position of the two preceding trials ($p(t_n|t_{n-1},t_{n-2}) = 1$) although the position of trial t_{n-1} alone is non-informative ($p(t_n|t_{n-1}) = .25$). In most studies published after Reed and Johnson (1994), the alternative sequence used in the transfer block also follows the SOC structure (but with a different surface expression) instead of the random sequence used previously. This allows for precise control of the statistical distribution of the alternative sequence and thus ensures that if a transfer effect is observed, it is due to the fact that the participants have learned the target sequence (see Reed & Johnson, 1994, for a justification of this manipulation).

Many parameters influencing sequence learning in the SRT task have been reported, such as sequence structure, alignment between stimulus locations and response key, or the presence of a secondary task (see Forkstam & Petersson, 2005; Schwarb & Schumacher, 2012; for reviews). However, the influence of yet another parameter, the duration of the Response-to-Stimulus Interval (RSI), remains unclear. Three hypotheses on the influence of RSI duration on sequence learning have been proposed in the literature. The first hypothesis states that the RSI would affect the amount of knowledge acquired about the sequence because of its impact on information processing in Working Memory (WM) (Frensch & Miner, 1994; Soetens et al., 2004). RSI duration would affect learning through the decay of stimulus representations in working memory : the shorter the duration of the RSI, the better the sequence would be learned. For the sake of clarity, we will refer to this hypothesis as the "Decay hypothesis". The second hypothesis states that the modulation of the RSI would not affect the learning of the sequence *per se*, but rather the performance in the SRT task (Willingham et al., 1997). According to this hypothesis, the duration of the RSI would affect the preparation of the response. Learning would be relatively equivalent across different RSI durations but could only be detected when the RSI is sufficiently short. With a long RSI, when the stimulus appears at an unexpected position in the transfer block, participants would have enough time to inhibit the learned response which would hide the transfer effect. We will refer to this hypothesis as the "Preparation hypothesis". Finally, the third hypothesis proposes that the RSI would mainly influence the ability to elaborate a conscious representation of the acquired knowledge (Cleeremans & Sarrazin, 2007; Destrebecqz & Cleeremans, 2001, 2003; Frensch & Miner, 1994; Huang et al., 2017; Kuhn & Dienes, 2006; Norman et al., 2007; Savalia et al., 2016; Soetens et al., 2004; W. B. Verwey & Wright, 2014; W. Verwey & Dronkers, 2019;

Willingham et al., 1997). The idea being that participants could only acquire explicit knowledge of the sequence when the duration of the RSI is long enough. When the RSI is very short or null, the knowledge acquired would remain largely implicit. Since the acquisition of item-based representations is limited to implicit learning, and acquiring more abstract knowledge requires explicit representation of the sequence, the duration of the RSI would affect the degree of abstraction of the acquired knowledge. We will refer to this hypothesis as the "Awareness hypothesis".

In what follows, we first review the empirical evidence for each of these hypotheses about the role of the RSI in the SRT task. To date, no consensus has been reached and all the aforementioned hypotheses remain plausible candidates. We then argue that the limitations of the method classically used to assess sequence learning (i.e., the transfer effect) make it unsuitable to settle between these hypotheses. We therefore assessed learning without using a transfer block. To achieve this, we took advantage from the specific properties of the sequence generated by the Fibonacci grammar which enable to measure learning continuously without having to compare performance to a transfer block.

Hypothesis 1: RSI duration affects sequence learning

The Decay hypothesis is based on the observation in some studies that the magnitude of the transfer effect decreases with increasing RSI duration. Frensch and Miner (1994) observed that the magnitude of the transfer effect was smaller for an RSI of 1500 ms than for an RSI of 500 ms. This detrimental effect of RSI lengthening on the transfer effect was replicated by other studies (Soetens et al., 2004; Stadler, 1995; Willingham et al., 1997). Frensch and Miner (1994) proposed that this detrimental effect would be due to the decay of the stimulus representation in WM. With a long RSI, the number of stimuli simultaneously active in WM would decrease, which would reduce the detection of sequence regularity and thus decrease the magnitude of the transfer effect. Nevertheless, the influence of the RSI on the magnitude of the transfer effect has not always been replicated, with multiple studies reporting no effect even with very different RSI values (Destrebecqz & Cleeremans, 2001, 2003; Huang et al., 2017; Norman et al., 2007) and none showing the opposite effect. For example, Destrebecqz and Cleeremans (2003) found that the magnitude of the transfer effect was identical for RSIs of 0 ms, 250 ms and 1500 ms. Moreover, although the claim that memory traces decay over time is assumed by multiple models of working memory (Barrouillet et al., 2004; Hommel, 1994; Mueller et al., 2003), it is not consensual due to the fact that effects typically attributed to decay can also actually be due to interference (Lewandowsky & Oberauer, 2009; Oberauer, 2019; Oberauer & Lewandowsky, 2013, 2014; Ricker et al., 2016).

Hypothesis 2: RSI duration would not affect sequence learning but response preparation

The Preparation hypothesis provides an alternative explanation for the finding of weaker transfer effects with long RSIs in the SRT task. Willingham et al. (1997) manipulated the duration of the RSI within participants. In the long-short condition, participants first performed four training blocks and a transfer block with a 1500 ms RSI. After this, participants performed another training block with the same target sequence but with an RSI reduced to 500 ms, followed by a final transfer block. In the short-long condition, the short RSI was initially presented, followed by the long RSI. Participants in the long-short condition showed no transfer effect when the RSI lasted 1500 ms. In contrast, the same participants displayed a transfer effect once the RSI duration was reduced to 500 ms, even though they performed only one training block at this presentation rate. To explain these results, Willingham et al (1997) suggested that sequence learning occurred when the RSI was 1500 ms long, but that the long presentation rate hid the learning effect, which was only visible when the RSI was short enough. The decrease in the magnitude of the transfer effect with increasing RSI duration frequently found in the literature would thus not reflect sequence learning but better response preparation. With a longer RSI, participants would have more time to prepare for the next trial, so the surprise effect induced by a stimulus arriving at an unexpected position would be less, resulting in a reduced transfer effect.

Another observation suggesting that RSI duration does not affect sequence learning itself comes from two studies that manipulated the duration of the RSI by using probabilistic target sequences (Norman et al., 2007; Shanks et al., 2003). In Normann et al. (2007), authors first created two SOC sequences, SOC-A and SOC-B. In order to make these sequences probabilistic, the authors simply manipulated the probability that the position of the stimulus at each trial was determined by the first or second sequence. During the training phase, the position of the stimulus was determined by the SOC-A sequence in 88% of the trials (high probability trials) and by the SOC-B sequence in 12% of the trials (low probability trials). The training phase was followed by a transfer block where the probabilities were reversed (i.e., in the transfer block, the probability that a trial was determined by the SOC-A sequence was .12 and by the SOC-B sequence was .88). The authors compared two conditions where the RSI lasted either 0 ms or 1000 ms. RSI duration did not affect the magnitude of the transfer effect, which was present at both 0 ms and 1000 ms. However, the results showed that in the training blocks, participants were slower for the low probability trials compared to the high probability trials and that this effect interacted with the RSI: the RTs difference between probable and improbable trials was bigger when the RSI lasted 0 ms than when it lasted 1000 ms. Similar results were reported by Shanks (2003). Thus, short RSIs exacerbated RTs differences between probable and improbable trials compared to long RSIs, although the transfer effect was identical in both cases. Normann et al. (2007) proposed that this effect could be due to a rapid shift of attention towards the next position predicted by the sequence. When participants have more time to prepare for the next trial (in the 1000 ms condition), they would be able to expand their attentional focus in anticipation of a target appearing anywhere. This broadening of attention would reduce the detrimental effect of a stimulus arriving at an unexpected position, resulting in smaller differences between probable and improbable trials compared to the 0 ms condition.

This hypothesis could also account for the common observation that longer RSIs give rise to faster RTs (Destrebecqz & Cleeremans, 2001, 2003; Frensch & Miner, 1994; Huang et al., 2017; Norman et al., 2006, 2007; Shanks et al., 2003; Soetens et al., 1985, 2004). As the duration of the RSI increases, participants have more time to prepare for the next trial, resulting in faster RTs. Some studies have also reported that the duration of the RSI affects the slopes of RTs in the training phase: the longer the RSI, the less steep the slopes (Destrebecqz & Cleeremans, 2003; Frensch & Miner, 1994; Soetens et al., 2004; Willingham et al., 1997). This may be due to the fact that when participants are slower (due to a short RSI), there is more room for RTs to decrease. Note however that the effect of RSI on slopes has not always been found (Destrebecqz & Cleeremans, 2001; Huang et al., 2017; Norman et al., 2006, 2007; Shanks et al., 2003). Therefore, RSI duration may not affect the learning of the sequence *per se*, but the preparation of the response in the SRT task.

Hypothesis 3: RSI duration affects sequence awareness

Many studies have put forward the hypothesis that the duration of the RSI would influence the ability to elaborate a conscious representation of the knowledge acquired in an SRT task (Cleeremans & Sarrazin, 2007; Destrebecqz & Cleeremans, 2001, 2003; Frensch & Miner, 1994; Huang et al., 2017; Kuhn & Dienes, 2006; Norman et al., 2007; Savalia et al., 2016; Soetens et al., 2004; Willingham et al., 1997). According to the Awarness hypothesis, participants could only acquire explicit knowledge of the sequence when the duration of the RSI is long enough. When the RSI is very short or absent, the knowledge acquired would remain largely implicit. It is assumed that implicit knowledge tends to be restricted to perceptual features of stimuli while explicit knowledge, on the other hand, is typically associated with more abstract representations (Cleeremans & Jiménez, 2002; Huang et al., 2017). As a consequence, RSI duration would affect the degree of abstraction of the acquired knowledge.

Destrebecqz and Cleeremans (2001, 2003) evaluated the influence of the duration of the RSI on sequence awareness by adding three tasks after the SRT task. In the *fragment recognition task*, participants had to judge if sequence fragments were identical or not to the target sequence. The underlying processes that allow success in this task, i.e., retrieving in memory the sequence learned during the SRT task and comparing it to the presented fragment, are seen as largely explicit (Perruchet et al., 1997; Perruchet & Amorim, 1992; Shanks et al., 2003; Shanks & Johnstone, 1999). In the *exclusion task*, participants had to produce a different sequence from the one previously learned. The exclusion task thus requires inhibiting the sequence learned, the underlying reasoning being that this inhibition can only take place if the representation of the sequence is explicit. Finally, in the *generation task*, participants had to reproduce the target sequence in a loop. Unlike the fragment recognition task and the exclusion task, success in the generation task does not require a conscious elaboration process and can be achieved on the basis of largely implicit knowledge (but note that the question of whether the generation task require only implicit knowledge has been debated, Goschke, 1998). Destrebecqz and Cleeremans (2001, 2003) observed poorer performance with a 0 ms RSI compared to longer RSIs in the two tasks that required explicit sequence knowledge, but no effect of RSI in the more implicit task. Although the studies by Destrebecqz and Cleeremans (2001, 2003) have been criticized for their lack of power (Wilkinson & Shanks, 2004), a number of other studies using alternative measures of awareness converge on the idea that RSI affects sequence awareness (see Forkstam & Petersson, 2005; Schwarb & Schumacher, 2012 for reviews).

If the duration of the RSI affects the degree of awareness of the sequence, then the degree of abstraction of the acquired knowledge should vary accordingly. The longer the RSI duration, the more abstract the acquired knowledge will be, as it provides more time for participants to process and analyze the sequence. Therefore, the learning of abstract structural rules should be impossible at short RSIs because the knowledge of the sequence is implicit. Huang et al., (2017) investigated whether the RSI influences the ability to acquire the underlying regularity of SOC sequences, which is that the position of the stimulus at trial *t* is determined by the position of the stimulus at trial t_{-1} and t_{-2}. To do so, they created two SOC sequences, SOC-A and SOC-B, which shared the same higher-order structural rule (i.e., they were both SOC sequences) but whose actual realization was different. In the first 10 blocks, participants were exposed to the SOC-A sequence. Block 11 was a transfer block where the order of the stimuli was randomized. In blocks 12, 13 and 14, the order of the stimuli followed the SOC-B sequence and block 15 was a second transfer block. The authors compared participants' performance when the RSI lasted 250 ms or 750 ms. The results showed that in the first transfer block, the transfer effect was equivalent for 250 ms and 750 ms RSIs conditions. However, in the second transfer block, only participants in the 750 ms RSI condition showed a transfer effect. Crucially, there was no transfer effect at either 250 ms or 750 ms in the control condition where participants were exposed to the SOC-B sequence for only 3 blocks, suggesting that the lack of transfer effect for SOC-B at 250 ms was not due to a lack of exposure. The authors' interpretation was that during the processing of the SOC-A sequence, participants in the 750 ms condition acquired the higher-order structural rule and reused it in the processing of the SOC-B sequence, whereas participants in the 250 ms condition failed to acquire the rule because the RSI was too short. These results are in line with the hypothesis that RSI does not affect learning as such, but the type of information that can be acquired.

Present study

In sum, both the Decay hypothesis and the Awareness hypothesis claim that RSI duration affects learning directly. According to the Decay hypothesis, RSI duration affects the amount of information that can be stored in working memory, such that shorter RSIs allow encoding more units from the sequence and therefore better sequence learning. According to the Awareness hypothesis, RSI duration affects the type of knowledge that can be acquired. Longer RSIs allow developing more abstract, higher-order properties of the sequence, and therefore better learning. In contrast to these two hypotheses, the Preparation hypothesis assumes that RSI duration does not affect sequence learning itself but participants' response preparation. Longer RSIs would reduce the surprise effect induced by a stimulus appearing at an unexpected position, thus hiding a potential transfer effect that would otherwise be observable at shorter RSIs.

Since the data reported in the literature on the impact of RSI on the transfer effect are sometimes inconsistent, and since there is no consensus on the interpretation of this impact, it may be relevant to switch to a different method to get fresh insight about the role of RSI in the SRT task. One important limitation of the transfer effect is that it is a relative measure of learning, since what is being measured is the participant's reaction to a change in the input. A slowdown in the transfer block is classically interpreted as due to the fact that the target sequence has been learned. However, because the transfer effect reflects the divergence between the target sequence and an alternative sequence, the slowdown entirely relies on the properties of this alternative sequence and how it differs from the target. Beyond the fact that this renders comparison among experiments difficult (since they vary on both the target and the alternative sequences), this method also fails to quantify *how much* participants have learned about the target sequence. Moreover, comparing averages of entire blocks does not take into account the fact that participants may continue to learn during the transfer blocks. It is therefore possible that intra-block learning hides the slowdown due to changes in the input : performance may be slower in the initial trials of the transfer block, and then improve such that the initial slowdown disappears in the average.

In the present study, we addressed these drawbacks of the transfer method. To do so, we took advantage of the properties of sequences generated by the Fibonacci grammar (Fib henceforth) which we used as a target sequence. The particularity of this sequence is that it allows us to quantify learning, trial by trial, without having to compare the performance of the participants to an alternative sequence. This provides us with a continuous measure of learning, without the need of transfer blocks. This is made possible by the fact that the regularities in the Fib sequences are dependent on one another: the learning of higher-order regularities is conditioned by the prior learning of lower order regularities. It is therefore possible to estimate how much participants have learned about the sequence by looking at the level at which these regularities have been extracted.

The Fib grammar is derived from the Lindenmayer formalism and was originally used to model algae growth (Lindenmayer, 1968; Vitányi & Walker, 1978). Recent studies have explored the processing of the Fib grammar in the SRT task (Schmid et al., 2023; Vender et al., 2019, 2020). These studies investigated whether participants process this sequence as a recursive nested structure of events. In particular, we observed in a previous study (Schmid et al., 2023) that the representation resulting from the processing of this sequence is similar to the natural constituent structure of the grammar. We proposed that in order to access this structure, the cognitive system would recursively merge the transitional probabilities between units of the sequence. This simple mechanism would result in a constituent structure similar to that of Fib because of the specific distribution of units in the sequence, which is aperiodic and self-similar. The Fib grammar is shown below and consists of two rewrite rules and contains a two-symbol alphabet:

$$0 \rightarrow 1$$

$$1 \rightarrow 01$$

The interpretation of these rules is the following: “0” is rewritten as “1” and “1” is rewritten as “01”. The successive application of these rules generates increasingly long sequences of 0s and 1s (henceforth refer as points). The name of this grammar comes from the fact that the number of points at each generation (i.e., each application of the rules) follows the Fibonacci sequence (Fig. 1C). This results in an asymmetry in the distribution of 0s and 1s that approximates the golden ratio (1.618) : in each generation, there are 1.618 times more 1s than 0s. Moreover, because the rewrite rules are recursive, each generation is the concatenation of the two previous ones. A sequence generated by the Fib grammar can therefore be parsed into smaller previous generations which are therefore the natural constituents of the grammar.

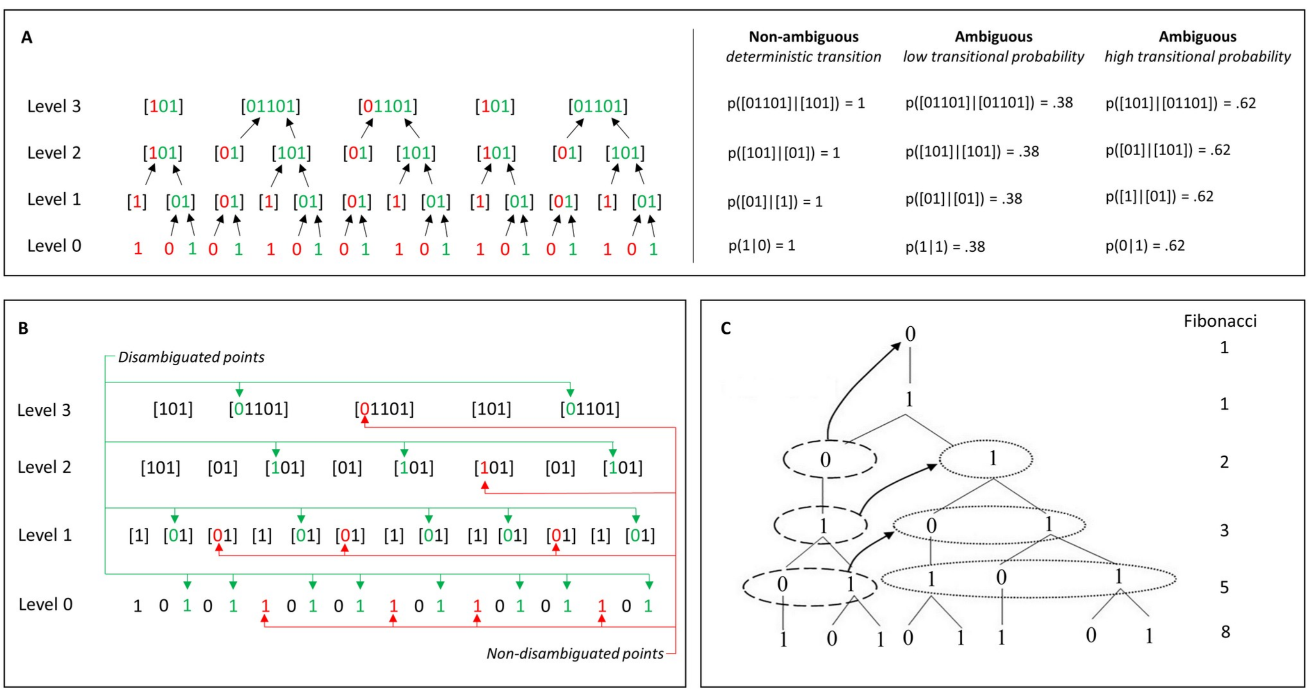

Fig. 1. (A) Left panel: depiction of the first three hierarchical levels of generation 7 of the Fibonacci grammar. Non-disambiguated points at each level are highlighted in red and disambiguated points in green. To form a new hierarchical level, points that span across a deterministic transition are combined together (this is illustrated by the arrows). The result is a new representation of the string that consists in the combination of points corresponding to natural higher-order constituents of the grammar (illustrated by the brackets). At each level, constituents spanning a deterministic transition can be combined to form an embedded hierarchy. Right panel: transition probabilities between constituents at each level. (B) Disambiguated points (green) and non-disambiguated points (red) for each hierarchical level for generation 7 of the Fibonacci grammar. In the present

study, we used generation 12 of the Fibonacci grammar that consists in 233 points. We did not illustrate this generation due space limitation, but the rationale is identical. (C) Derivation of the Fibonacci grammar for the first 5 generations. The right column shows the number of symbols at each generation, which maps the Fibonacci sequence. Arrows and circles highlight the hierarchical constituency of the grammar.

The interest of the sequences generated by this grammar is that they are aperiodic and self-similar. In the classical SRT task, a SOC sequence is presented in a loop, so the sequence the participant is exposed to is periodic. Thus, learning the sequence theoretically allows to predict *all* future trials with *certainty*. This is not possible in sequences generated by the Fib grammar because of their aperiodicity: it is impossible to predict *all* the trials with *certainty* because there is no pattern that repeats in a loop. However, these sequences are not random: they present regularities, but the distribution of these regularities is aperiodic. If we examine the first order transitional probabilities (i.e., the conditional probability of a point according to the point preceding it) of these sequences, we see that three transitions are possible (Fig. 1A right panel). The first transition is deterministic: a 0 is always followed by a 1 ($p(1|0)=1$). The two other transitions are probabilistic: a 1 is followed by a 0 in 62% of the cases ($p(0|1)= .62$) and by a 1 in 38% of the cases ($p(1|1)= .38$). Points that follow a first order deterministic transition (i.e., 1s that appear after a 0) can be predicted with certainty on the basis of what precedes them, whereas this is not the case for points that follow a probabilistic transition. However, sequences generated by the Fib grammar are also self-similar, which means that the transitional probabilities between points are found also in the transitions between constituents. This implies that some of the points that follow a probabilistic transition can appear in a constituent that follows a deterministic transition. Thus, accessing to these higher-order deterministic transitions allows to predict with certainty a subset of points that follow a lower-order probabilistic transition (Fig. 1A left panel).

In a previous study (Schmid et al., 2023), we proposed that in order to access the constituent structure of the grammar, the cognitive system would start by merging the points linked by a deterministic transition. This would result in the creation of a constituent on the basis of which the cognitive system could further detect new higher-order deterministic transitions (i.e., a deterministic transitions between constituents).

Let's take an example: at the surface level, 0 is always followed by 1; the merge of these points gives rise to the constituent [01]. This results in a new representation (which we call a new hierarchical level) where the sequence is partitioned into two constituents: [1] and [01]. At this level, constituent [01] can be followed either by constituent [1] (p([1]|[01])=.62) or by constituent [01] (p([01]|[01])=.38), while constituent [1] is always followed by constituent [01] (p([01]|[1])=1). The cognitive system can merge again the constituents that span across a deterministic transition, which results in the creation of the constituent [101]. The new representation of the sequence then consists of two new constituents: [01] and [101]. The sequence being self-similar, the transition between these constituents is identical to that of lower levels, and merging of the deterministic transition p([101]|[01]) = 1 would lead to the creation of a new hierarchical level. The key property to understand is that the first point of a constituent that follows a deterministic transition at level n always follows a probabilistic transition at level *n-1*. Thus, a point that is ambiguous at level *n* can be disambiguated if it appears at level *n+1* in a constituent that follows a deterministic transition. For example, the first point of the constituent [01] (i.e., the 0) at level 1 always follows a probabilistic transition at level 0 (p(0|1)=.62). If the cognitive system has detected the higher-order deterministic transition p([01]|[1]) = 1, then a subset of the points that follow a probabilistic transition at level 0 (i.e., the 0s that appear in constituent [01] when it follows constituent [1]) can now be predicted with certainty (Fig. 1B). For clarity, we will call *disambiguated points* the points that follow a (higher-order) deterministic transition and *non-disambiguated* points the points that follow a (higher-order) probabilistic transition at the same hierarchical level. To test the hypothesis that participants develop a hierarchical structure based on the recursive merge of deterministic transitions, we implemented sequences of the Fib grammar in an SRT task where 0s and 1s were transformed into red and blue circles and presented in the center of a screen (Schmid et al., 2023). The RSI lasted 500 ms. The results showed a greater decrease in RTs for disambiguated points than for non-disambiguated points at levels 0, 1, 2 and 3, suggesting that participants had reached the 3rd hierarchical level.

In the present study, we asked to what extent RSI duration impacts learning in the Fib grammar. Because of the self-similar character of the sequences, the number of hierarchical levels is theoretically infinite[1], thus, there is no a priori limitation in the amount of knowledge that participants can acquire. The use of the Fib grammar therefore makes it possible to evaluate the depth of learning without having to use the transfer method.

In order to systematically explore the influence of the RSI in the SRT task, we conducted three experiments where we manipulated the duration of the RSI. The RSI lasted 1000 ms in Experiment 1, 250 ms in Experiment 2 and 100 ms in Experiment 3. In all experiments, we used the same paradigm as Schmid et al. (2023) where sequences of the Fib grammar were implemented in the SRT task. The 0s and 1s were transformed into red or blue circles (respectively) and presented sequentially in the center of a screen. Participants had to press the button associated with the displayed color. The answer made the circle disappear and triggered the next trial.

Each hypothesis makes distinct predictions about how RSI duration will affect the height of the hierarchical structure elaborated by the participants. According to the Decay hypothesis, the WM representation of the points/constituents deteriorates over time; this should make it more difficult to

1 Note that the hierarchical depth can of course only be infinite for an infinite string. In the present study, the shortest sequences were 144 points long and potentially involved up to 11 hierarchical levels, which is likely well beyond the processing capacity of the cognitive system.

merge deterministic transitions. Therefore, the height of the hierarchical structure built by the participants should increase with the shortening of the RSI. According to the Awareness hypothesis, the duration of the RSI affects the type of knowledge that can be acquired. Under the common assumption in this field that the construction of abstract, hierarchical structure requires explicit knowledge, the height of the hierarchical structure should decrease with the shortening of the RSI. The predictions of the Awareness hypothesis regarding the height of the hierarchical structure are thus the opposite of those of the Decay hypothesis. Finally, according to the Preparation hypothesis, RSI duration does not affect learning per se but the preparation of the response in the SRT task. Thus, the height of the hierarchical structure should not vary with RSI duration. However, the length of the RSI should still affect the results because it affects the time to prepare for the next trial. Therefore, participants should be faster on average the longer the RSI. This should go along with less steep RTs slopes because if participants are faster at long RSIs, there is less room for improvement.

In order to test if RSI duration affects hierarchical learning, we conducted a first analysis in which we evaluated the height of the hierarchical structure in each experiment in the same way as in our previous study (Schmid et al., 2023). Hierarchical elaboration generates expectations about the structure of the input, which the participants' RTs reflect (Huettel et al., 2002; Hyman, 1953; Lynn et al., 2020; McCarthy & Donchin, 1981; Sternberg, 1969). Hierarchical learning therefore manifests in terms of steeper slopes of RTs for disambiguated points at a given level compared to the slopes of non-disambiguated points at the same level. To control for asymmetry in the distribution of 0s and 1s in the sequence (i.e., 1s are more frequent than 0s), we compared at each level only 1s to 1s and 0s to 0s. Since the Preparation hypothesis predicts that RSI duration should not affect the height of the hierarchical structure but the time to prepare for the next trial, we also conducted a second analysis where we compared the average RTs and slopes of each experiment. An important point to clarify is that the Preparation hypothesis makes predictions about the average RTs and slopes for all points regardless of their ambiguity status. That’s because RSI duration is a constant that affects all trials in the same way. We therefore considered disambiguated and non-disambiguated points of all levels jointly in this analysis. In order to have a wider range of RSI durations, we also included in this analysis the results of Schmid et al. (2023) with a 500 ms RSI.

Anticipating the results, the first analysis revealed that the duration of the RSI affected the height of the hierarchical structure in a non-linear way: participants reached the 3rd hierarchical level when the RSI lasted 250 ms, whereas they only reached the 2nd hierarchical level for 100 ms and 1000 ms RSIs. None of the three hypotheses can account for this U-shape pattern of results. The second analysis partially met the predictions of the Preparation hypothesis. Concerning the average RTs, participants were slower with a 100 ms RSI compared to all other RSI durations, however, there was no difference between RSIs of 250 ms, 500 ms, 1000 ms. Concerning the average slopes, we found again a non-linear, U-shape effect of RSI duration: the slopes for the 250 ms and 500 ms RSIs were both steeper than the slopes for the 1000 ms and 100 ms RSIs, which did not differ. This result cannot be explained by the Preparation hypothesis.

In summary, the continuous measure of performance used in the present study combined with the testing of multiple RSIs show that that there is actually an optimal time window for learning in the SRT task. None of the hypotheses proposed in the literature can capture the non-linear effect of RSI duration on both the height of the hierarchical structure represented and the evolution of RTs through the task. We propose that this effect is due to an information compression mechanism determined by the interaction between the encoding capacity of the cognitive system (which corresponds to the amount of entropy the system can encode per unit of time) and the source rate of information transmission (the amount of entropy per unit of time sent by the source) (Radulescu et al., 2019, 2021; Shannon, 1948).

Experiment 1 : 1000 ms RSI

In Experiment 1, the RSI lasted 1000 ms. To avoid that the RSI of 1000 ms exhausted the participants, we made small design change in Experiment 1 compare to Schmid et al. (2023) design (where the RSI lasted 500 ms) : we presented more blocks (7 instead of 5) but shorter (144 trials per block instead of 233). As a result, the overall number of trials per participant slightly differs from Schmid et al. (2023) (1008 trials instead of 1165). Deidentified data collected in Experiment 1 are posted at https://osf.io/pfgbu/?view_only=39add1c9dffe4b0b82c748e6574a73a8

Methods

Participants

One hundred and eighty participants (49 men and 131 women; mean age 24 years old) recruited through announcements at the University of Geneva participated in the experiment. Thirty-two participated as volunteers and the remaining 148 were paid 10 CHF. All participants reported normal or corrected-to-normal vision.

Materials

The training sequence was composed of two elements and had a length of 50. The order was pseudo-randomized and elements had the same frequency. The training sequence included multiple non-grammatical sub-sequences such as 00 or 111. The longest Fib-grammatical sub-sequence had a length of 6. In the experimental blocks, the sequence consisted of generation 11 of the Fibonacci grammar which has 144 points. Each experimental block corresponded to a full generation.

Design and procedure

Each trial consisted of a red or blue circle 100px in diameter presented at the center of the screen which correspond, respectively, to 0 and 1 in a string generated by the Fib grammar. The circles disappeared after the response of the participant, or after 1200 ms, if no response was given. The RSI lasted 1000 ms. Participants responded by pressing the button corresponding to the color of the circle. Participants responded using the X and M keys of a QWERTZ keyboard (X=blue, M=red). The experiment started with a training block of 50 trials that was identical for all the participants. After the training block, participants did 7 experimental blocks of 144 trials. Instructions were displayed on the screen and participants had to click on a button to start the experiment. Participants were instructed to respond as quickly as possible. Pre-testing showed that the error rate in the task was extremely low, which is not surprising given the simplicity of the task, so the emphasis on speed alone was intended to increase the error rate and avoid ceiling effects. No information related to the grammar was given. Between each block, a message was displayed saying that participant had to press the key “enter” to start the next block, participants were told at the beginning of the experiment that they could take as much time as they wanted between each block. Stimuli were presented electronically using the E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). The experiment was conducted using a desktop computer running on windows 7 with a 17’’ inch screen with a 1280*1020 pixels resolution. The computer screen was placed approximately 60cm from the participants. The experiment lasted approximately 30 minutes.

Data analyses

We removed six participants who had an error rate superior to 3 *SD* to the mean error rate in at least one block. Reaction times and accuracy were both modelled as dependent variables. We removed from the analysis all the trials where participants did not respond after 1200 ms (321 trials). For the analysis of reaction times, only trials with a correct answer were included. Homoscedasticity and normality were checked by visual inspection of residual plots. Data from the remaining 174 participants were analyzed with linear mixed-effects models as implemented in the lme4 package for R (Bates et al., 2014; R Core Team, 2022). Models included two fixed-effect factors and their interaction: *Exposure*, *Ambiguity*, and *Exposure*Ambiguity*. Exposure was treated as a continuous variable with a value of 0 for trials of the 1st experimental block, and of 1, 2, 3, 4, 5 and 6 for trials of the 2nd, 3d, 4th , 5th 6th and 7th blocks. This factor being continuous, it allowed us to have only one estimate which represents the evolution (i.e., the slope) of performance throughout the experiment across all participants. Ambiguity is a discrete variable contrasting disambiguated and non-disambiguated points and operationalized differently depending on the level at which its effect is explored (it is labeled *Ambiguity* $level_n$ according to the level at which it has been operationalized). We entered as fixed effects the factors *Ambiguity* $level_n$ (Disambiguated vs. Non-disambiguated), *Exposure*, and the interaction *Exposure*Ambiguity*. The modality "Non-disambiguated" of the factor *Ambiguity* $level_n$ was always set as the intercept of the models. As random effects, the models had intercepts for *Participants*. P-values were calculated by way of the Satterthwaites's approximation to degrees of freedom with the lmerTest package (Kuznetsova et al., 2015).

In order to explore the height of the hierarchical structure elaborated by the participants, we tested the effect of Ambiguity starting at level 0. We stopped the analysis as soon as the effect was no longer significant. In the present experiment, we conducted the analysis at levels 0, 1, 2 and 3.

Results

Processing of Level 0

Analyses of reaction times showed a main effect of *Exposure* (β = -11.94, SE = 0.17, t = -69.06, p < .000) with a mean reduction of reaction times of 72 ms from block 1 to block 7. There was also a main effect of *Ambiguity* $level_0$ (β = -50.23, SE = 0.72, t = -69.64, p < .000) with disambiguated points being faster than non-disambiguated ones by 50 ms. The interaction *Ambiguity* $level_0$* *Exposure* was also significant (β = -9.59, SE = 0.36, t = -26.73, p < .000) with a more important reduction over time for disambiguated points ($M_{block7 - block1}$ = -95 ms) than non-disambiguated points ($M_{block7 - block1}$ = -34 ms) ($M_{block7 - block1}$ indicates the mean difference between blocks 7 and 1). Results are shown in Fig. 2.

Concerning accuracy, we found a main effect of *Exposure* (β = -0.07, *SE* = 0.01, *z* = -7.303, *p*< .000) with a mean reduction of accuracy of 1 % from block 1 to block 7. There was also a main effect of *Ambiguity* $level_0$ (β = 2.18, *SE* = 0.05, *z* = 47.849, *p*< .000) with higher accuracy for disambiguated points (*M* = 0.99) than for non-disambiguated points (*M* = 0.93). The effect of *Exposure* significantly interacted with *Ambiguity* $level_0$ (β = 0.27, *SE* = 0.02, *z* = 11.66, *p*< .000) with accuracy increasing for disambiguated points over exposure ($M_{block7-block1}$ = 0.01) and decreasing for non-disambiguated points ($M_{block7-block1}$ = -0.04). Results are shown in Table 1.

Processing of Level 1

Analyses of reaction times showed a main effect of *Exposure* (β = -9.64, *SE* = 0.22, *t* = -44.05, *p* < .000) with a mean reduction of reaction times of 58 ms from block 1 to block 7. There was also a main effect of *Ambiguity* $level_1$ (β = -35.82, *SE* = 0.90, *t* = -39.62, *p* < .000) with disambiguated points being faster than non-disambiguated ones by 36 ms. The interaction *Ambiguity* $level_1$* *Exposure* was also significant (β = -7.05, *SE* = 0.45, *t* = -15.65, *p* < .000) with a more important reduction over time for disambiguated points ($M_{block7-block1}$ = -73 ms) than non-disambiguated points ($M_{block7-block1}$ = -31 ms). Results are shown in Fig. 2.

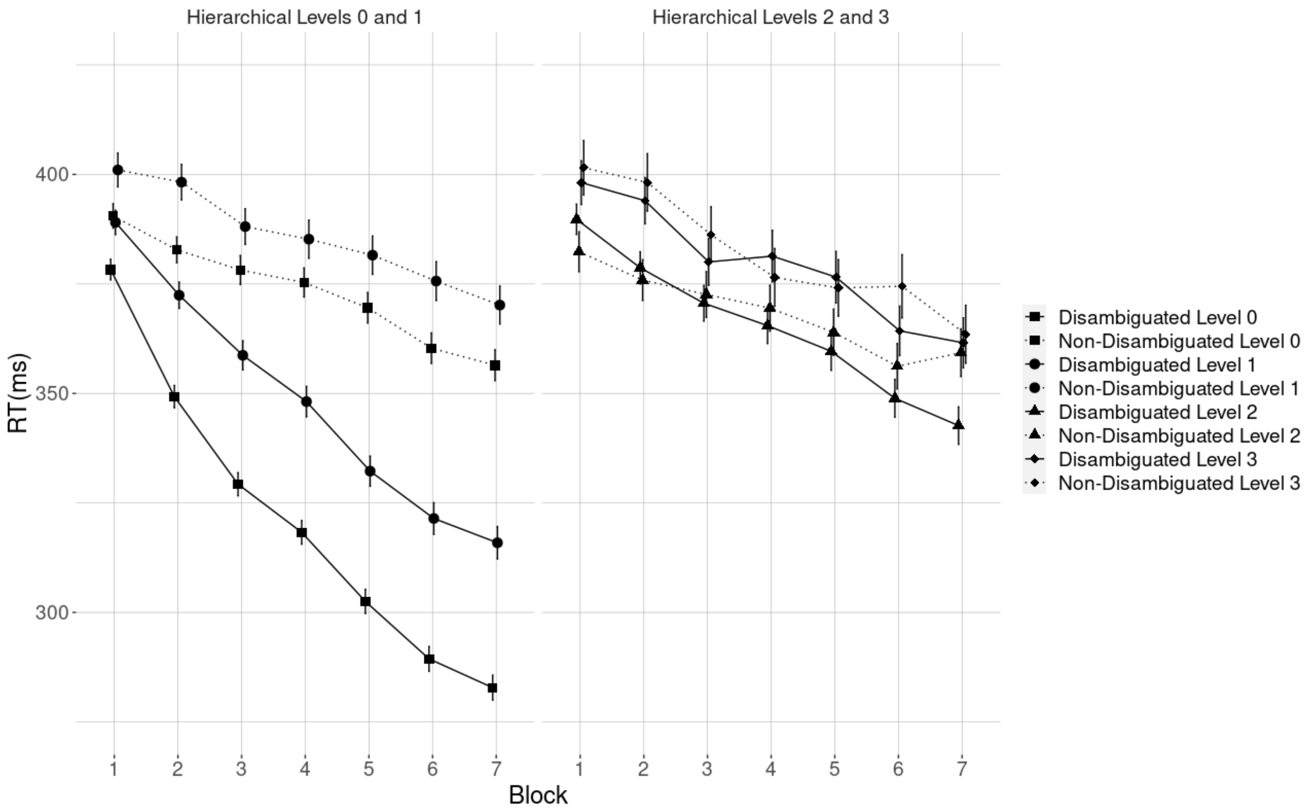

Fig. 2. *Mean RT (ms) for Disambiguated and Non-disambiguated points for Hierarchical Levels 0 and 1 (left) and for Hierarchical Levels 2 and 3 (right) by Block in Experiment 1. Errors bars denote the 95% confidence interval.*

Concerning accuracy, we found a main effect of *Exposure* (β = -0.09, *SE* = 0.01, *z* = -9.196, *p*< .000) with a mean reduction of accuracy of 2.4 % from block 1 to block 7. There was also a main effect of *Ambiguity level*$_1$ (β = 1.06, *SE* = 0.04, *z* = 27.059, *p*< .000) with accuracy higher for disambiguated points (*M* = 0.97) than for non-disambiguated points (*M* = 0.93). The effect of *Exposure* significantly interacted with *Ambiguity level*$_1$ (β = 0.10, *SE* = 0.02, *z* = 5.005, *p*< .000) with accuracy increasing for disambiguated points over time ($M_{block7\,-\,block1}$ = 0.003) and decreasing for non-disambiguated points ($M_{block7\,-\,block1}$ = -0.06). Results are shown in Table 1.

Table 1

Mean Proportion (M) and Standard Deviation (SD) of Correct Responses for Disambiguated and Non-Disambiguated Points by Hierarchical Levels and Blocks in Experiment 1.

		Block 1		Block 2		Block 3		Block 4		Block 5		Block 6		Block 7	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Level 0	Disambiguated	0.98	0.13	0.99	0.11	0.99	0.08	0.99	0.08	0.99	0.09	0.99	0.08	0.99	0.08
	Non-disambiguated	0.96	0.20	0.94	0.23	0.93	0.25	0.92	0.26	0.92	0.28	0.91	0.28	0.92	0.28
Level 1	Disambiguated	0.97	0.16	0.98	0.15	0.97	0.16	0.97	0.18	0.97	0.17	0.97	0.17	0.97	0.16
	Non-disambiguated	0.96	0.20	0.94	0.23	0.93	0.26	0.92	0.27	0.92	0.27	0.91	0.28	0.90	0.30
Level 2	Disambiguated	0.95	0.21	0.94	0.23	0.94	0.25	0.93	0.26	0.92	0.27	0.92	0.28	0.92	0.27
	Non-disambiguated	0.97	0.19	0.94	0.23	0.93	0.25	0.92	0.28	0.91	0.28	0.91	0.29	0.91	0.28
Level 3	Disambiguated	0.96	0.20	0.94	0.24	0.92	0.27	0.92	0.27	0.92	0.27	0.92	0.28	0.90	0.30
	Non-disambiguated	0.96	0.19	0.95	0.22	0.93	0.25	0.93	0.26	0.92	0.27	0.92	0.28	0.91	0.29

Processing of Level 2

Analyses of reaction times showed a main effect of *Exposure* (β = -5.73, *SE* = 0.27, t = -20.893, p < .000) with a mean reduction of reaction times of 34 ms from block 1 to block 7. There was also a main effect of *Ambiguity level*$_2$ (β = -5.07, *SE* = 1.14, t = -4.436, p < .000) with disambiguated points being faster than non-disambiguated ones by 5 ms. The interaction *Ambiguity level*$_2$**Exposure* was also significant (β = -3.85, *SE* = 0.57, t = -6.762, p < .000) with a more important reduction over exposure for disambiguated points ($M_{block7 – block1}$ = -47 ms) than non-disambiguated points ($M_{block7 – block1}$ = -23 ms). Results are shown in Fig. 2.

Concerning accuracy, we found a main effect of *Exposure* (β = -0.12, *SE* = 0.01, z = -11.66, p< .000) with a mean reduction of accuracy of 4 % from block 1 to block 7. There was no effect of *Ambiguity level*$_2$ (β = 0.05, *SE* = 0.04, z = 1.33, p = .184) and the interaction *Ambiguity level*$_2$**Exposure* was not significant (β = 0.04, *SE* = 0.02, z = 1.852, p= .064). Results are shown in Table 1.

Processing of Level 3

Analyses of reaction times showed a main effect of *Exposure* (β = -5.22, *SE* = 0.37, *t* = -14.23, *p* < .000) with a mean reduction of reaction times of 31 ms from block 1 to block 7. There was no effect of *Ambiguity level$_3$* (β = -1.94, *SE* = 1.50, *t* = -1.293, *p* = .196). The interaction *Ambiguity level$_3$*Exposure* was also not significant (β = -0.03, *SE* = 0.7, *t* = -0.042, *p* = .966). Results are shown in Fig. 2.

Concerning accuracy, we found a main effect of *Exposure* (β = -0.12, *SE* = 0.01, *z* = -9.864, *p*< .000) with a mean reduction of accuracy of 6 % from block 1 to block 7. There was also a main effect of *Ambiguity level$_3$* (β = -0.13, *SE* = 0.05, *z* = -2.622, *p* = .009) with accuracy higher for non-disambiguated points (*M* = 0.92) than for disambiguated points (*M* = 0.93). The interaction *Ambiguity level$_3$* Exposure* was however not significant (β = 0.02, *SE* = 0.03, *z* = 0.848, *p*= .396). Results are shown in Table 1.

Discussion

In Experiment 1, we found that RTs of disambiguated points at levels 0, 1 and 2 decreased significantly more through exposure than their non-disambiguated counterparts. We also found that accuracy tended to decreased over time at levels 0, 1, 2 and 3. However, this decrease was modulated by Ambiguity : at levels 0 and 1, accuracy for non-disambiguated points decreased through exposure while it increased for disambiguated points. At level 2, the decrease in accuracy was identical for disambiguated and non-disambiguated points. Thus, even if this decrease in accuracy suggests a speed-accuracy trade-off, it cannot explain the difference in RTs between disambiguated and non-disambiguated points. This decrease in accuracy was also observed in our previous study (Schmid et al., 2022) and could be due to instructions that emphasized speed of response or to the boredom induced by the long RSI or the simplicity of the task. At level 3, RTs did not differ between disambiguated and non-disambiguated points. There was however a main effect in accuracy with a better performance for disambiguated points compared to non-disambiguated points. However, this effect did not interact with time, suggesting that the effect was present from the beginning and does therefore not reflect learning. Taken together, these results suggest that with an RSI of 1000 ms, participants reached the second hierarchical level.

Experiment 2 : 250 ms RSI

In Experiment 2, the RSI lasted 250 ms. Based on pilot testing, we noticed that the fatigue induced by this duration is reduced compared to that induced by the RSI of 1000 ms in Experiment 1. We therefore reproduced the design of Schmid et al. (2023) with 5 blocks of 233 trials each. As a result, the total number of trials per participant is slightly higher in Experiment 2 (1165 trials) than in Experiment 1 (1008 trials). Apart from these minor adjustments, the design of Experiment 2 was identical to that of Experiment 1. Deidentified data collected in Experiment 2 are posted at https://osf.io/pfgbu/?view_only=39add1c9dffe4b0b82c748e6574a73a8

Methods

Participants

One hundred and fifty participants (men and women; mean age years old) recruited using Prolific (www.prolific.co) participated in the experiment. Participants were paid 3.75 £. All participants reported normal or corrected-to-normal vision.

Materials

The training sequence was identical to that of experiment 1. However, instead of generation 11 of the Fib grammar we used generation 12 which contain 233 points.

Design and procedure

The procedure was identical to that of Experiment 1 except for the following elements. During the training block, when the participants made an error, the experiment stopped and a message appeared to remind them the color – key association, the experiment resumed after 3000 ms. In the experimental blocks, no message appeared when they made an error. After the training block, participants did 5 experimental blocks of 233 trials. The experiment was created using PsychoPy (Peirce et al., 2019) and conducted online on the website Pavlovia (www.pavlovia.org). Participants were asked to perform the experiment in a quiet environment where they could not be disturbed. The experiment lasted approximately 20 minutes.

Data analyses

We removed one participant who had a number of timeout trials 3 SD above the mean answered trials. We also removed four participants who had an error rate superior to 3 *SD* to the mean error rate in at least one block. We removed from the analysis all the trials where participants did not respond after 1200 ms (578 trials). Data from the remaining 145 participants were analyzed in the same way as in Experiment 1.

Results

Processing of Level 0

Analyses of reaction times showed a main effect of *Exposure* (β = -19.81, *SE* = 0.24, t = -80.99, p < .000) with a mean reduction of reaction times of 87 ms from block 1 to block 5. There was also a main effect of *Ambiguity level*$_0$ (β = -51.24, *SE* = 0.71, t = -71.30, p < .000) with disambiguated points being faster than non-disambiguated ones by 51 ms. The interaction *Ambiguity level*$_0$* *Exposure* was also significant (β = -14.98, *SE* = 0.51, t = -29.64, p < .000) with a more important reduction over exposure for disambiguated points ($M_{block5 – block1}$ = -110 ms) than non-disambiguated points ($M_{block5 – block1}$ = -45 ms) ($M_{block5 – block1}$ indicates the mean difference between blocks 1 and 5). Results are shown in Fig. 3.

Concerning accuracy, there was a main effect of *Exposure* (β = 0.04, *SE* = 0.01, z = -2.634, p = .008) with a mean reduction of accuracy of 0.4 % from block 1 to block 5. There was also a main effect of

Ambiguity level$_0$ (β = 1.99, *SE* = 0.05, *z* = 42.59, *p* < .000) with higher accuracy for disambiguated points (*M* = 0.99) than for non-disambiguated points (*M* = 0.94). The interaction *Exposure*Ambiguity level*$_0$ was significant (β = 0.14, *SE* = 0.03, *z* = 4.294, *p* < .000) with accuracy increasing over exposure for disambiguated points ($M_{block5 - block1}$ = 0.004) and decreasing for non-disambiguated points ($M_{block5 - block1}$ = -0.02). Results are shown in Table 2.

Processing of Level 1

Analyses of reaction times showed a main effect of *Exposure* (β = -19.26, *SE* = 0.34, *t* = -57.22, *p* < .000) with a mean reduction of reaction times of 77 ms from block 1 to block 5. There was also a main effect of *Ambiguity level*$_1$ (β = -87.57, *SE* = 1.00, *t* = -87.86, *p* < .000) with disambiguated points being faster than non-disambiguated ones by 88 ms. The interaction *Ambiguity level*$_1$** Exposure* was also significant (β = -18.40, *SE* = 0.70, *t* = -26.36, *p* < .000) with a more important reduction over exposure for disambiguated points ($M_{block5 - block1}$ = -108 ms) than non-disambiguated points ($M_{block5 - block1}$ = -30 ms). Results are shown in Fig. 3.

Concerning accuracy, we found a main effect of *Exposure* (β = -0.12, *SE* = 0.01, *z* = -9.492, *p* < .000) with a mean reduction of accuracy of 2.4 % from block 1 to block 5. There was also a main effect of *Ambiguity level*$_1$ (β = 1.56, *SE* = 0.04, *z* = 39.80, *p*< .000) with accuracy higher for disambiguated points (*M* = 0.98) than for non-disambiguated points (*M* = 0.90). The effect of *Exposure* significantly interacted with *Ambiguity level*$_1$ (β = 0.11, *SE* = 0.03, *z* = 3.803, *p*< .000) with accuracy increasing over exposure for disambiguated points ($M_{block5 - block1}$ = 0.004) and decreasing for non-disambiguated points ($M_{block1 - block75}$ = -0.06). Results are shown in Table 2.

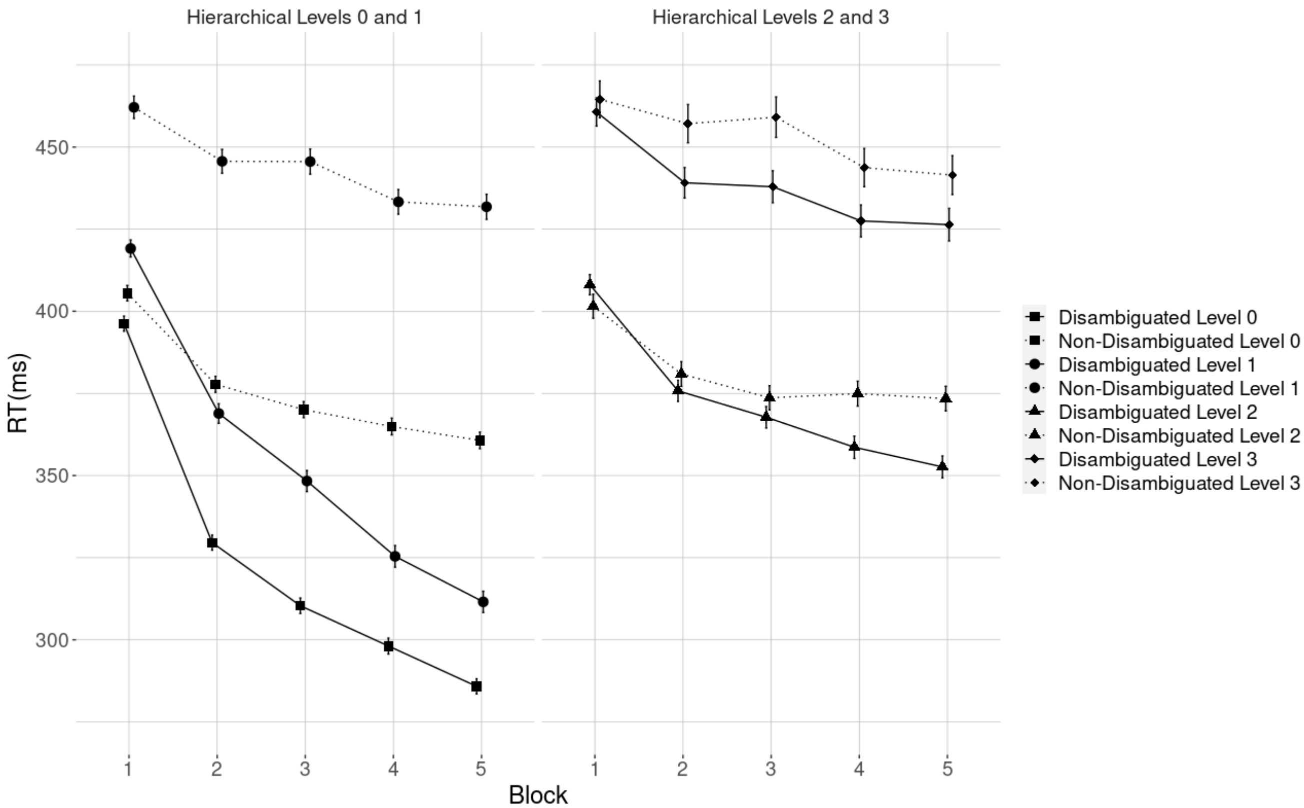

Fig. 3. *Mean RT (ms) for Disambiguated and Non-disambiguated points for Hierarchical Levels 0 and 1 (left) and Hierarchical Levels 2 and 3 (right) by Block in Experiment 2. Errors bars denote the 95% confidence interval.*

Processing of Level 2

Analyses of reaction times showed a main effect of *Exposure* (β = -10.32, *SE* = 0.34, t = -30.19, $p <$ .000) with a mean reduction of reaction times of 41 ms from block 1 to block 5. There was also a main effect of *Ambiguity level*$_2$ (β = -7.16, *SE* = 0.99, t = -7.21, $p <$ .000) with disambiguated points being faster than non-disambiguated ones by 7 ms. The interaction *Ambiguity level*$_2$* *Exposure* was also significant (β = -6.33, *SE* = 0.70, t = -9.039, $p <$ .000) with a more important reduction over exposure for disambiguated points ($M_{block5 - block1}$ = -55 ms) than non-disambiguated points ($M_{block5 - block1}$ = -28 ms). Results are shown in Fig. 3.

Concerning accuracy, there was a significant effect of *Exposure* (β = -0.07, *SE* = 0.02, *z* = -4.316, *p* < .000) with a mean augmentation of accuracy of 1.6 % from block 1 to block 5. There was also a main effect of *Ambiguity level*$_2$ (β = -0.09, *SE* = 0.04, *z* = -2.103, *p* = .035) with accuracy higher for non-disambiguated points (*M* = 0.942) than for disambiguated points (*M* = 0.937). The effect of *Exposure* significantly interacted with *Ambiguity level*$_2$ (β = 0.14, *SE* = 0.03, *z* = 4.508, *p*< .000) with accuracy decreasing more over exposure for non-disambiguated points ($M_{block5 - block1}$ = -0.03) than for disambiguated points ($M_{block5 - block1}$ = -0.005). Results are shown in Table 2.

Table 2

Mean Proportion (M) and Standard Deviation (SD) of Correct Responses for Disambiguated and Non-Disambiguated Points by Hierarchical Levels and Blocks in Experiment 2.

		Block 1		Block 2		Block 3		Block 4		Block 5	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Level 0	Disambiguated	0.98	0.11	0.99	0.09	0.99	0.09	0.99	0.09	0.99	0.09
	Non-disambiguated	0.95	0.21	0.94	0.24	0.94	0.24	0.93	0.25	0.93	0.25
Level 1	Disambiguated	0.98	0.15	0.98	0.15	0.97	0.16	0.97	0.16	0.98	0.16
	Non-disambiguated	0.94	0.24	0.91	0.29	0.88	0.32	0.88	0.32	0.88	0.32
Level 2	Disambiguated	0.94	0.23	0.93	0.25	0.94	0.25	0.94	0.24	0.94	0.24
	Non-disambiguated	0.97	0.18	0.94	0.23	0.94	0.24	0.93	0.25	0.93	0.25
Level 3	Disambiguated	0.93	0.26	0.91	0.29	0.88	0.32	0.89	0.31	0.89	0.31
	Non-disambiguated	0.95	0.22	0.90	0.30	0.88	0.33	0.87	0.34	0.87	0.33

Processing of Level 3

Analyses of reaction times showed a main effect of *Exposure* (β = -7.34, *SE* = 0.52, *t* = -14.01, *p* < .000) with a mean reduction of reaction times of 30 ms from block 1 to block 5. There was also a main effect of *Ambiguity level*$_3$ (β = -14.38, *SE* = 1.55, *t* = -9.277, *p* < .000) with disambiguated points being faster than non-disambiguated ones by 14 ms. The interaction *Ambiguity* $level_3$* *Exposure* was significant (β = -2.14, *SE* = 1.09, *t* = -1.967, *p* = .049) with a more important reduction over exposure for disambiguated points ($M_{block5 - block1}$ = -34 ms) than non-disambiguated points ($M_{block5 - block1}$ = -23 ms). Results are shown in Fig. 3.

Concerning accuracy, we found a main effect of *Exposure* (β = -0.15, *SE* = 0.02, *z* = -9.951, *p*< .000) with a mean reduction of accuracy of 5.6 % from block 1 to block 5. There was no main effect of *Ambiguity* $level_3$ (β = 0.06, *SE* = 0.05, *z* = 1.296, *p* = .195). The interaction *Exposure** *Ambiguity* $level_3$ was however significant (β = 0.09, *SE* = 0.03, *z* = 2.762, *p* = .006) with accuracy decreasing less over exposure for disambiguated points ($M_{block5 - block1}$ = -0.04) than for non-disambiguated points ($M_{block5 - block1}$ = -0.08). Results are shown in Table 2.

Processing of level 4

Analyses of reaction times showed a main effect of *Exposure* (β = -6.21, *SE* = 0.53, *t* = -11.716, *p* < .000) with a mean reduction of reaction times of 25 ms from block 1 to block 5. There was no effect of *Ambiguity* $level_4$ (β = -0.79, *SE* = 1.53, *t* = -0.515, *p* = .607). The interaction *Ambiguity* $level_4$* *Exposure* did not reach significance (β = -1.04, *SE* = 1.08, *t* = -0.962, *p* = .336).

Concerning accuracy, we found a main effect of *Exposure* (β = -0.15, *SE* = 0.03, *z* = -5.856, *p*< .000) with a mean reduction of accuracy of 3.4 % from block 1 to block 5. There was a main effect of *Ambiguity* $level_4$ (β = 0.15, *SE* = 0.07, *z* = 2.090, *p* = .037) with accuracy higher for disambiguated points (*M* = 0.946) than for non-disambiguated points (*M* = 0.939). The interaction *Exposure** *Ambiguity* $level_4$ was however not significant (β = 0.01, *SE* = 0.05, *z* = 0.134, *p* = .893).

Discussion

The results of Experiment 2 show that RTs of disambiguated points decreased significantly more through exposure than their non-disambiguated counterparts at hierarchical levels 0, 1, 2 and 3. As in

Experiment 1, accuracy tended to decrease throughout the experiment, but this decrease was modulated by ambiguity: at levels 0 and 1, accuracy increased for disambiguated points, while it decreased for their non-disambiguated counterparts. At levels 2 and 3, accuracy decreased for both disambiguated and non-disambiguated points, but the decrease was significantly greater for non-disambiguated points. Finally, there was no effect at level 4 in the RTs. Accuracy turned out to be higher for non-disambiguated points than for disambiguated points, but this effect did not evolve over time and has a small effect size (less than 1%); we will therefore not interpret this effect. In summary, results of Experiment 2 suggest that participants reached the third hierarchical level when the RSI lasted 250 ms.

Experiment 3: 100 ms RSI

In Experiment 3, the RSI was 100 ms long. We took advantage of the fact that this short RSI reduces the total time to complete the experiment to increase the number of trials per participant. Thus, Experiment 3 was slightly longer but the number of trials was more than twice as large. This results in a higher statistical power compared to Experiments 1 and 2, however, as will become clear, this increase in statistical power did not favor the appearance of significant effects. Participants were exposed to 7 blocks of 377 trials (generation 13 of the Fib grammar) for a total of 2639 trials (compared to 1008 trials for experiment 1 and 1165 trials for experiment 2). Apart from these differences, the design of Experiment 3 was identical to that of Experiments 1 and 2. Deidentified data collected in Experiment 3 are posted at https://osf.io/pfgbu/?view_only=39add1c9dffe4b0b82c748e6574a73a8

Methods

Participants

One hundred participants (44 men and 56 women; mean age 24.4 years old) recruited using Prolific (www.prolific.co) participated in the experiment. Participants were paid 3.75 £. All participants reported normal or corrected-to-normal vision.

Materials

The training sequence was identical to that of experiment 1 and 2. However, we used generation 13 the Fib grammar which contain 377 points instead of generation 11 used in Experiment 1 and generation 12 used in Experiment 2.

Design and procedure

The procedure was identical to that of Experiment 2 except that participants did 7 experimental blocks of 377 trials. The experiment was created using PsychoPy (Peirce et al., 2019) and conducted online on the website Pavlovia (www.pavlovia.org). Participants were asked to perform the experiment in a quiet environment where they could not be disturbed. The experiment lasted approximately 30 minutes.

Data analyses

One participant was removed for not providing answers in two blocks. We also removed five participants who had an error rate superior to 3 *SD* to the mean error rate in at least one block. We removed from the analysis all the trials where participants did not respond after 1200 ms (1375 trials). For the analysis of reaction times, only trials with a correct answer were included. Homoscedasticity and normality were checked by visual inspection of residual plots. Data from the remaining 94 participants were analyzed in the same way as in Experiment 1 and 2.

Results

Processing of Level 0

Analyses of reaction times showed a main effect of *Exposure* (β = -11.47, *SE* = 0.17, *t* = -69.38, *p* < .000) with a mean reduction of reaction times of 69 ms from block 1 to block 7. There was also a main effect of *Ambiguity* $level_0$ (β = -53.67, *SE* = 0.68, *t* = -78.39, *p* < .000) with disambiguated points being faster than non-disambiguated ones by 54 ms. The interaction *Ambiguity* $level_0$* *Exposure* was also significant (β = -5.23, *SE* = 0.34, *t* = -15.28, *p* < .000) with a more important reduction over exposure for disambiguated points ($M_{block7 – block1}$ = -91 ms) than non-disambiguated points ($M_{block7 – block1}$ = -52 ms) ($M_{block7 – block1}$ indicates the mean difference between blocks 1 and 7). Results are shown in Fig. 4.

Concerning accuracy, there was no effect of *Exposure* (β = 0.01, *SE* = 0.01, *z* = 1.54, *p* = .123). There was however a main effect of *Ambiguity* $level_0$ (β = 1.57, *SE* = 0.04, *z* = 42.637, *p* < .000) with higher accuracy for disambiguated points (*M* = 0.99) than for non-disambiguated points (*M* = 0.95). The interaction *Exposure*Ambiguity* $level_0$ was not significant (β = -0.02, *SE* = 0.02, *z* = -0.977, *p* = .328). Results are shown in Table 3.

Processing of Level 1

Analyses of reaction times showed a main effect of *Exposure* (β = -13.30, *SE* = 0.22, *t* = -61.07, *p* < .000) with a mean reduction of reaction times of 80 ms from block 1 to block 7. There was also a main effect of *Ambiguity* $level_1$ (β = -141.87, *SE* = 0.91, *t* = -155.89, *p* < .000) with disambiguated points being faster than non-disambiguated ones by 142 ms. The interaction *Ambiguity* $level_1$* *Exposure* was also significant (β = -13.38, *SE* = 0.45, *t* = -29.66, *p* < .000) with a more important reduction over exposure for disambiguated points ($M_{block7 – block1}$ = -117 ms) than non-disambiguated points ($M_{block7 – block1}$ = -31 ms). Results are shown in Fig. 4.

Concerning accuracy, we found a main effect of *Exposure* (β = -0.09, *SE* = 0.01, *z* = -11.89, *p*< .000) with a mean reduction of accuracy of 2.8 % from block 1 to block 7. There was also a main effect of

Ambiguity level₁ (β = 1.67, *SE* = 0.03, *z* = 50.89, *p*< .000) with accuracy higher for disambiguated points (*M* = 0.98) than for non-disambiguated points (*M* = 0.89). The effect of *Exposure* significantly interacted with *Ambiguity level₁* (β = 0.07, *SE* = 0.02, *z* = 4.431, *p*< .000) with accuracy increasing over exposure for disambiguated points ($M_{block7 - block1}$ = 0.004) and decreasing for non-disambiguated points ($M_{block7 - block1}$ = -0.07). Results are shown in Table 3.

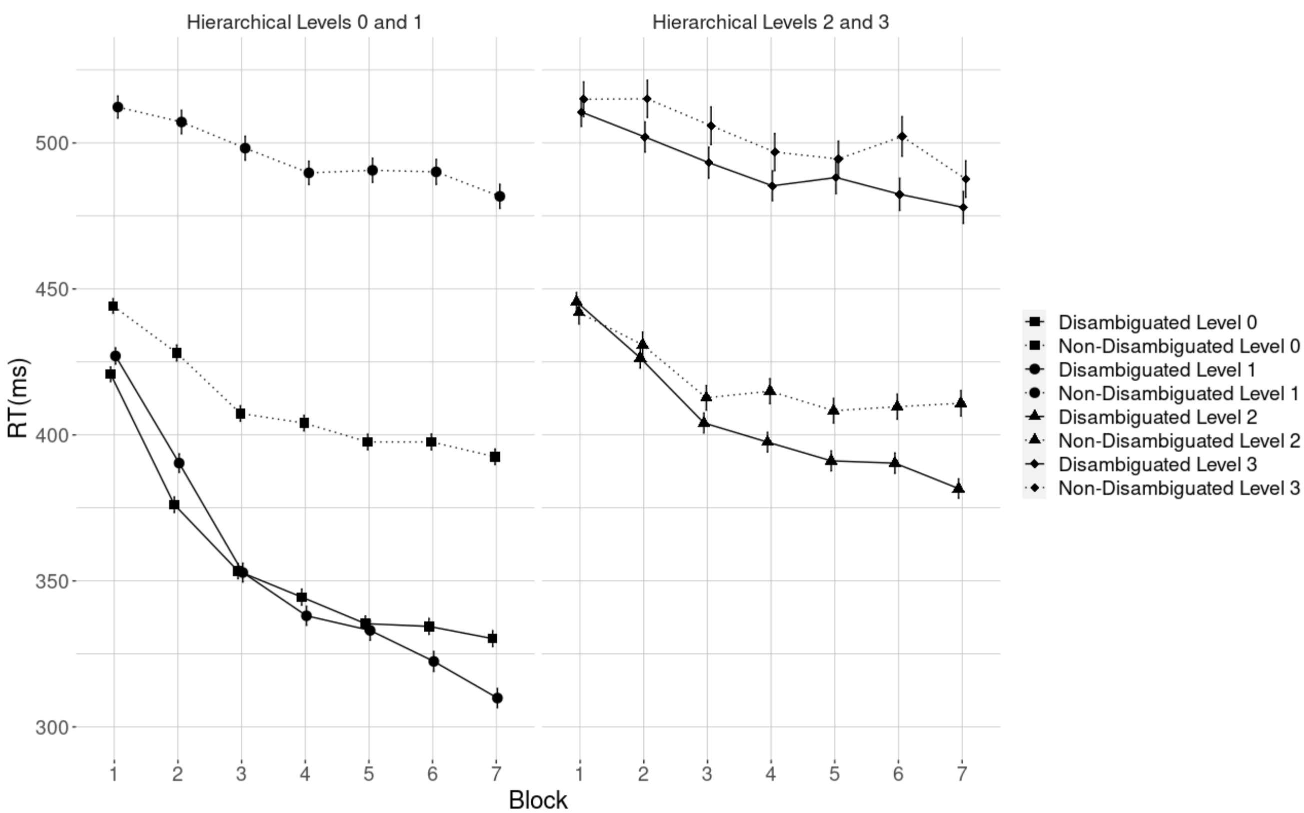

Fig. 4. *Mean RT (ms) for Disambiguated and Non-disambiguated points of Hierarchical Levels 0 and 1 (left) and for Hierarchical Levels 2 and 3 (right) by Block in Experiment 3. Errors bars denote the 95% confidence interval.*

Table 3

Mean Proportion (M) and Standard Deviation (SD) of Correct Responses for Disambiguated and Non-Disambiguated Points by Hierarchical Levels and Blocks in Experiment 3.

		Block 1		Block 2		Block 3		Block 4		Block 5		Block 6		Block 7	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Level 0	Disambiguated	0.99	0.10	0.99	0.10	0.99	0.11	0.99	0.10	0.98	0.12	0.99	0.9	0.99	0.10
	Non-disambiguated	0.95	0.22	0.94	0.23	0.95	0.22	0.96	0.21	0.95	0.21	0.95	0.22	0.95	0.21
Level 1	Disambiguated	0.98	0.14	0.98	0.15	0.98	0.15	0.98	0.15	0.97	0.17	0.98	0.15	0.97	0.16
	Non-disambiguated	0.94	0.23	0.91	0.29	0.88	0.32	0.89	0.32	0.87	0.33	0.87	0.33	0.87	0.33
Level 2	Disambiguated	0.95	0.23	0.94	0.24	0.95	0.23	0.96	0.21	0.95	0.22	0.95	0.22	0.96	0.20
	Non-disambiguated	0.96	0.19	0.95	0.21	0.96	0.21	0.96	0.21	0.96	0.20	0.95	0.22	0.94	0.22
Level 3	Disambiguated	0.94	0.24	0.90	0.30	0.88	0.33	0.90	0.31	0.87	0.34	0.88	0.33	0.88	0.33
	Non-disambiguated	0.95	0.22	0.92	0.27	0.89	0.31	0.87	0.33	0.87	0.33	0.87	0.34	0.88	0.33

Processing of Level 2

Analyses of reaction times showed a main effect of *Exposure* (β = -8.10, *SE* = 0.23, t = -35.56, p < .000) with a mean reduction of reaction times of 48 ms from block 1 to block 7. There was also a main effect of *Ambiguity* $level_2$ (β = -12.55, *SE* = 0.94, t = -13.36, p < .000) with disambiguated points being faster than non-disambiguated ones by 13 ms. The interaction *Ambiguity* $level_2$* *Exposure* was also significant (β = -4.86, *SE* = 0.47, t = -10.359, p < .000) with a more important reduction over exposure for disambiguated points ($M_{block7 – block1}$ = -64 ms) than non-disambiguated points ($M_{block7 – block1}$ = -31 ms). Results are shown in Fig. 4.

Concerning accuracy, there was no effect of *Exposure* (β = 0.02, *SE* = 0.01, z = 1.805, p = .071). There was a main effect of *Ambiguity* $level_2$ (β = -0.13, *SE* = 0.04, z = -3.136, p = .002) with accuracy higher for non-disambiguated points (M = 0.96) than for disambiguated points (M = 0.95). The effect of

Exposure significantly interacted with *Ambiguity* $level_2$ (β = 0.09, *SE* = 0.02, *z* = 4.324, *p*< .000) with accuracy increasing for disambiguated points over exposure ($M_{block7 - block1}$ = 0.01) and decreasing for non-disambiguated points ($M_{block7 - block1}$ = -0.02). Results are shown in Table 3.

Processing of Level 3

Analyses of reaction times showed a main effect of *Exposure* (β = -4.75, *SE* = 0.35, *t* = -13.449, *p* < .000) with a mean reduction of reaction times of 29 ms from block 1 to block 7. There was also a main effect of *Ambiguity* $level_3$ (β = -10.33, *SE* = 1.46, *t* = -7.091, *p* < .000) with disambiguated points being faster than non-disambiguated ones by 10 ms. The interaction *Ambiguity* $level_3$* *Exposure* was however not significant (β = -0.86, *SE* = 0.72, *t* = -1.182, *p* = .237). Results are shown in Fig. 4.

Concerning accuracy, we found a main effect of *Exposure* (β = -0.11, *SE* = 0.01, *z* = -12.31, *p*< .000) with a mean reduction of accuracy of 6.7 % from block 1 to block 7. There was no effect of *Ambiguity* $level_3$ (β = -0.04, *SE* = 0.04, *z* = -0.986, *p* = .324). The interaction *Exposure** *Ambiguity* $level_3$ was also not significant (β = 0.03, *SE* = 0.02, *z* = 1.792, *p* = .073). Results are shown in Table 3.

Discussion

The results of Experiment 3 showed that the RTs for disambiguated points decreased more through exposure than their non-disambiguated counterparts at levels 0, 1 and 2. There was no effect in accuracy at level 0. At levels 1 and 2, accuracy increased through exposure for disambiguated points and decreased for non-disambiguated points. At level 3, there was only a main effect on RTs with disambiguated points processed faster than non-disambiguated points. However, the interaction was non-significant, suggesting that this effect do not reflect learning. Taken together, these results suggest that participants reached the second hierarchical level when the RSI lasted 100 ms. It is interesting to note that even though the number of trials was significantly higher in this experiment, this greater exposure did not improve learning compared to Experiment 2 where participants reached level 3. This is in line with a recent meta-analysis (Isbilen & Christiansen, 2022), which reported that the amount of exposure does not influence learning. It may be that the amount of exposure plays only a minor role above a certain threshold.

Comparison of 1000 ms, 500 ms, 250 ms and 100 ms RSIs

In this second analysis, we test the predictions of the Preparation hypothesis that RSI duration affects participants' control of the response. With longer RSIs, participants have more time to prepare their responses which should result in faster RTs than for shorter RSI. Conversely, the slope of RTs should be steeper the shorter the RSI because there is more room for improvement. Since the RSI is the same throughout the trials, the influence of RSI duration on preparation is also constant and is therefore not expected to vary between the different types of points. We therefore compared the average RTs and slopes without distinguishing between disambiguated and non-disambiguated points at the different levels. In order to have a wider range of RSI duration, we also integrated in this analysis our previous results where the RSI lasted 500 ms (Schmid et al., 2023). Except for the duration of the RSI, the design of this experiment was strictly identical to that of Experiment 2 where the RSI lasted 250 ms. Since the amount of exposure varied in each experiment, we considered only the first 1008 trials of each experiment in order to have the same number of trials in each experiment. This number corresponds to the number of trials in Experiment 1, which was the shortest.

Methods

Materials

Since Experiment 1 contains 1008 experimental trials and is the shortest, we considered in the analyses the first 1008 experimental trials of experiments 1 (RSI = 1000 ms), 2 (RSI = 250 ms) and 3 (RSI = 100 ms). We also included in the analysis the first 1008 experimental trials of Schmid et al, (2023) where the RSI was 500 ms. We did not include in the analysis the trials of the training block where the order of the points was random.

Data analysis

Models included two fixed-effect factors and their interaction: *Exposure, RSI,* and *Exposure *RSI*. Exposure was treated as a continuous variable with a value of 0 for the first trial and 1007 for the last trial. This factor being continuous, it allowed us to have only one estimate which represents the evolution (i.e., the slope) of RTs throughout the experiments across all participants. Trials where an incorrect answer was given were not included in the analysis. RSI is a between subject discrete variable contrasting RSI duration. We entered as fixed effects the factors *RSI* (1000 ms vs 500 ms vs 250 ms vs 100 ms), *Exposure,* and the interaction *Exposure*RSI*. The modality “100 ms” of the factor *RSI* was set as the intercept of the models. As random effects, the models had intercepts for *Participants*. Since the factor RSI contained 4 modalities and that the Preparation hypothesis makes predictions on all comparisons, we had to run the model 3 times. In order to control for type 1 error, we applied the Bonferroni correction for multiple testing by dividing the alpha level by 3. We therefore considered as significant the p-value lower than .01667. P-values were calculated by way of the Satterthwaites’s approximation to degrees of freedom with the lmerTest package (Kuznetsova et al., 2015).

Results

Results showed a main effect of the *Exposure* (β = -0.089, SE = 0.0005, t = -160.91, p < .000) indicating that RTs decreased across exposure. RTs in the "100 ms" condition (M = 416 ms; SD = 146 ms) were significantly slower than those of the "250 ms" condition (M = 365 ms; SD = 138 ms) (β = -51.87, SE = 8.39, t = -6.182, p < .000), “500 ms” condition (M = 363 ms; SD = 136 ms) (β = -53.54, SE = 8.24, t = -6.494, p < .000) and “1000 ms” condition (M = 350 ms; SD = 134 ms) (β = -67.40, SE = 8.11, t = -8.310, p < .000). The mean RTs of the "1000 ms" condition did not differ from those of the

"500 ms" condition (β = 13.86, *SE* = 6.95, *t* = 1.994, *p* = .047) and "250 ms" condition (β = 15.53, *SE* = 7.12, *t* = 2.180, *p* = .030). Finally, there was no difference between the "500 ms" and "250 ms" conditions (β = 1.66, *SE* = 7.27, *t* = 0.229, *p* = .819). Concerning the interaction *RSI*Exposure*, there was no difference in slope of RTs between the "100 ms" condition and "1000 ms" condition (β = -0.03, *SE* = 0.001, *t* = -1.622, *p* = .105). There was also no difference between the "250 ms" and "500 ms" conditions (β = 0.002, *SE* = 0.002, *t* = 1.179, *p* = .238). However, the RTs decreased more through exposure for the "250 ms" condition compare to the "1000 ms" condition (β = -0.02, *SE* = 0.001, *t* = -13.872, *p* < .000) and the "100 ms" condition (β = -0.02, *SE* = 0.001, *t* = -10.209, *p* < .000). Finally, RTs decreased more through exposure for the "500 ms" condition compare to the "1000 ms" condition (β = -0.02, *SE* = 0.001, *t* = -15.396, *p* < .000) and the "100 ms" condition (β = -0.02, *SE* = 0.002, *t* = -11.404, *p* < .000). Results are shown in Fig. 5.

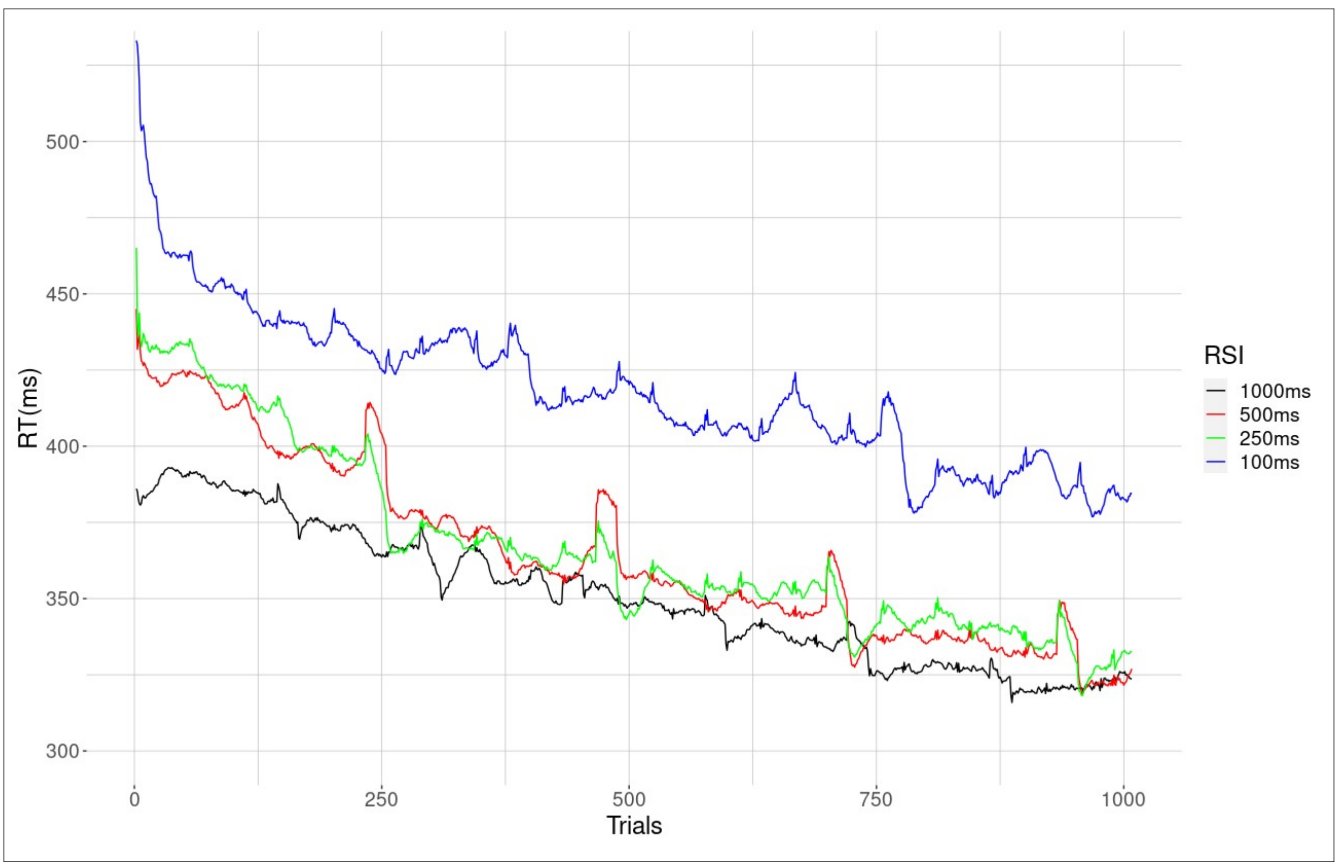

Fig. 5. *Moving average of RTs for the first 1008 experimental trials as a function of RSI duration. The moving average is calculated over the 21 trials that precede trial t. We removed trial 1 of each experiment from the graph because it was processed extremely slowly and was out of the frame.*

Discussion

According to the Preparation hypothesis, average RTs should increase as the RSI is reduced. In line with this prediction, participants were indeed slower when the RSI was 100 ms long compared to all other conditions. However, we found no difference in average RTs between the 1000 ms, 500 ms and 250 ms RSIs, contrary to the predictions of the Preparation hypothesis. We doubt that this lack of effect is due to a lack of power given the number of participants (572) and the number of trials per participant (1008). We also found that the duration of the RSI affected the slope of the RTs in a non-linear way: RTs decreased more strongly for RSIs of 250 ms and 500 ms compared to RSIs of 1000 ms and 100 ms. Again, this result cannot be explained by the Preparation hypothesis which predicted a linear relationship between slope steepness and RSI.

General discussion

Three hypotheses have been put forward to explain the role of RSI duration on sequence learning in the SRT task. According to the Decay hypothesis (Frensch & Miner, 1994; Soetens et al., 2004), the duration of the RSI affects information processing in WM. As the duration of the RSI increases, the trace of the stimuli would tend to decrease, which would decrease the number of stimuli simultaneously active in WM. As a result, the detection of sequence regularities would become more difficult as RSI increases and conversely, learning would be better with shorter RSIs. According to the Preparation hypothesis (Norman et al., 2007; Shanks et al., 2003; Willingham et al., 1997), the duration of the RSI does not affect sequence learning as such but the preparation of the response; sequence learning would be relatively equivalent across different RSI values. According to the Awareness hypothesis, the duration of RSI affects the development of explicit knowledge of the target sequence, and therefore the learning of the structural rules underlying the sequence (Cleeremans & Sarrazin, 2007; Destrebecqz & Cleeremans, 2001, 2003; Frensch & Miner, 1994; Huang et al., 2017; Kuhn & Dienes, 2006; Norman et al., 2007; Savalia et al., 2016; Soetens et al., 2004; Willingham et al., 1997). Learning becomes increasingly explicit with increasing RSI duration and remain largely implicit at short RSI duration. Consequently, knowledge requiring conscious elaboration such as structural rules could only be acquired when the RSI is sufficiently long.

In the present study, we tested the predictions of each of these hypotheses by implementing sequences generated by the Fib grammar in the SRT task. The use of these sequences allowed us to quantify learning during processing without having to expose participants to alternative sequences. We manipulated the duration of the RSI in three experiments: 1000 ms in Experiment 1, 250 ms in Experiment 2 and 100 ms in Experiment 3. In all three experiments, RTs for disambiguated points decreased through exposure more than RTs for non-disambiguated points at levels 0, 1 and 2. At level 3, this effect was only present in Experiment 2 where the RSI lasted 250 ms. In all experiments, accuracy decreased systematically for non-disambiguated points. In contrast, accuracy for disambiguated points either increased through exposure or decreased, but to a lesser extent than non-disambiguated points. These results suggest that participants built a hierarchical structure up to the second hierarchical level when the RSI lasted 1000 ms and 100 ms and reached the third hierarchical level when the RSI lasted 250 ms. Taken together with our previous finding that participants also reached the third level with an RSI of 500 ms (Schmid et al., 2023), it seems that the duration of RSI has a non-linear effect on sequence learning. This non-linear effect of RSI duration on the height of the hierarchical structure cannot be explained by any of the hypotheses put forward.

If the RSI affects sequence learning through stimulus decay in WM, then merging points/constituents should be more difficult as the RSI is longer. Therefore, the height of the hierarchical structure should increase with shortening of the RSI. The Decay hypothesis cannot explain the fact that the hierarchical structure elaborated by participants in Experiment 3 (100 ms RSI) is lower than in Experiment 2 (250 ms RSI) and in Schmid et al. (2023) (500 ms RSI). Furthermore, exposure was more than twice as long when the RSI lasted 100 ms, this higher exposure should have strengthened the trace of stimuli in WM, and thus promoted hierarchical elaboration.

According to the Preparation hypothesis, RSI duration do not affect sequence learning but the preparation of the response, therefore the height of the hierarchical structure built by the participants (i.e. that reflect sequence learning) should be identical across different RSI duration. The non-linear effect of RSI duration on learning we observed cannot be accounted for by this hypothesis. However, according to this hypothesis, RSI duration affect the preparation of the responses as it modulate the time to prepare for the next trial. This should result in participants being faster overall the longer the RSI and, conversely, the overall slope of RTs should be steeper the shorter the RSI. To test this second prediction, we compared the average RTs and the overall slope of decrease of RTs of the three experiments. We also included in this analysis the results of our previous study where the RSI lasted 500 ms (Schmid et al., 2023). We only took into account the first 1008 trials in order to have the same number of trials in each experiment (1008 corresponds to the number of trials in Experiment 1 which was the shortest). The results showed that participants were significantly slower when the RSI was 100 ms long compared to all other experiments. There was no difference between the 1000 ms, 500 ms and 250 ms RSIs after the p-values have been corrected for multiple testing. We also found that the decrease in RTs across exposure was greater when the RSI lasted 500 ms and 250 ms compared to when it lasted 1000 ms and 100 ms. There was no difference in slope between the 500 ms and 250 ms RSIs and between the 1000 ms and 100 ms RSIs. These results cannot be explained by the Preparation hypothesis which predicts that the effect of RSI duration on slopes is linear. These results also address a potential confound regarding the way we assessed learning. Indeed, we considered that reaching a hierarchical level results in a larger decrease in RTs for disambiguated points compared to non-disambiguated points of the same level. To the extent that the Preparation hypothesis also predicts an effect on slopes, it could be that the effect of the RSI on the height of the hierarchical structure is in fact due to the preparation of the response by the participants. However, this explanation cannot account for the fact that the overall slopes of the RTs as well as the height of the hierarchical structure built by the participants were non-linearly affected by the duration of the RSI.

Finally, according to the Awareness hypothesis, the length of the RSI would affect the type of knowledge that can be acquired. If hierarchical learning in the Fib grammar involves higher-order structural rules, then the height of the hierarchical structure should increase with the lenghtening of the RSI. Our results can be interpreted in two different ways, depending on the assumption retained. If the Awareness hypothesis is true, i.e. if RSI duration affects the implicit aspect of learning and that structural rules can only be acquired explicitly, since RSI duration affected learning in a non-linear way, our results mean that Fib grammar processing does not require conscious elaboration of structural rules. Note that this is not in contradiction with the hypothesis that the Fib grammar gives rise to hierarchical elaboration. Indeed, we do not claim that the participants have learned the rewriting rules of the Fib grammar in order to access its hierarchical structure, nor that the knowledge they have developed is akin to abstract structural rules. Our hypothesis is that participants build a hierarchical structure from the input by recursively merging points/constituents that span across a deterministic transition. Because the sequence generated by the Fib grammar are aperiodic and self-similar, this mechanism results in a hierarchical structure similar to the natural constituent structure of the Fib grammar. The hierarchical elaboration is thus driven by the particular distributional regularities of Fib and not by the fact that the participants would have acquired the underlying rules of the grammar. If one assumes that hierarchical building in Fib necessarily requires the elaboration structural rules, either participants have acquired these rules implicitly, or RSI duration does not affect the implicit/explicit aspect of learning in the SRT task. However, our results cannot disentangle between these hypothesis since we did not assess participants' awareness of Fib. Note that even if this had been the case, it is not certain that checking participants' consciousness could have provided useful information. Indeed, this evaluation is indirect and is done by adding additional tasks after the learning phase. These tasks therefore measure what remains in memory after processing and not the degree of awareness of the mechanisms involved during encoding.

In summary, none of the hypotheses put forward can explain the non-linear effect of RSI duration on performance. The first question raised by these results is why a non-linear effect of RSI has, to our knowledge, never been reported in the literature. One possible explanation could be that this is due to sampling bias. If, as is often the case, only two RSI durations are compared, then the non-linear effect of RSI is invisible. Our observation would simply come from the fact that we compared more than two RSI durations. While this non-linear effect of the RSI may never have been observed due to sampling bias in RSI duration, this still does not explain its existence.

In the following, we interpret these findings through the lens of recent applications of Shannon's information theory to sequence processing (Pothos, 2010; Radulescu et al., 2019, 2021; Shannon, 1948). According to Shannon information theory (1949), if the amount of information in a signal exceeds the encoding capacity of the receiver, another encoding method should be used to limit the loss of information. Changing the encoding method means compressing the input signal into another format. According to the *Information Premise* (Pothos, 2010), the cognitive system would tends to represents new information with as little uncertainty as possible. To accomplish this, the input is recoded (i.e. compressed) in a way that minimizes the entropy of the system's representational state.

According to Radulescu et al. (2019, 2021), acquisition of higher order knowledge result from the tension between the amount of information contained in the input and the cognitive system's limited encoding power in term of memory and processing speed. The encoding power of the cognitive system is defined as the amount of information per unit of time that it is able to process. When the signal exceeds the encoding power, it can be encoded with minimal loss as long as a sufficiently efficient compression method is available. If there is no compression method suitable for the amount of information in the input, the loss of information will increase. Thus, increasing the volume of information per unit of time delivered to the cognitive system compels it to compress the input into a more abstract format as long as the system has a sufficiently powerful compression method at its disposal. If the amount of information delivered to the system exceeds the most efficient compression method, this will result in a loss of information.

In what follows, we adopt this proposal to explain the non-linear effect of RSI duration. We consider that hierarchical elaboration is the manifestation of information compression. Thus, the height of the hierarchical structure elaborated by the participants reflects the degree of compression of the sequence. RSI duration determines the volume of information per unit of time that is delivered to the participants. This hypothesis explain the non-linear effect of RSI duration in the following way: when the RSI lasts 1000 ms, the amount of information to encode per unit of time would not require compressing the sequence beyond the second hierarchical level. As the RSI shortens, the amount of information per unit of time increases and the sequence is further compressed, thus explaining why participants reach the third hierarchical level with RSIs of 500 ms and 250 ms. When the RSI lasts 100 ms, the volume of information is too large to compress the sequence without loss, which explains why participants reached only the second hierarchical level.

Conclusion

The results of the present study broadly replicate previous observations that participants extract a hierarchical structure when processing Fib grammar sequence in the SRT task (Schmid et al., 2023; Vender et al., 2019, 2020). This adds to the growing interest in Fibonacci grammar processing (Geambaşu et al., 2016, 2020; Krivochen et al., 2018; Shirley, 2014). Our results also suggest that there is an optimal temporal window for sequence learning in the SRT task. It is possible that these results stem from a tension between the limited encoding power of the cognitive system and the amount of information per unit of time delivered to the system. An open question is whether this non-linear effect of the RSI is specific to the sequence generated by the Fib grammar or whether it can be replicated in other types of sequences. Future work is therefore necessary.

References

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *Lme4 : Linear mixed-effects models using Eigen and S4. R package version 1.1-7*.

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. *Implicit Learning and Consciousness, 2002*, 1–40.

Cleeremans, A., & Sarrazin, J.-C. (2007). Time, action, and consciousness. *Human Movement Science*, *26*(2), 180-202. https://doi.org/10.1016/j.humov.2007.01.009

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Sciences*, *26*(9), 751–766. https://doi.org/10.1016/j.tics.2022.06.010

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

Destrebecqz, A., & Cleeremans, A. (2003). Temporal effects in sequence learning. In L. Jiménez (Éd.), *Advances in Consciousness Research* (Vol. 48, p. 181-213). John Benjamins Publishing Company. https://doi.org/10.1075/aicr.48.11des

Forkstam, C., & Petersson, K. M. (2005). Towards an explicit account of implicit learning. *Current Opinion in Neurology*, *18*(4), 435-441. https://doi.org/10.1097/01.wco.0000171951.82995.c4

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, *22*(1), 95-110. https://doi.org/10.3758/BF03202765

Geambaşu, A., Ravignani, A., & Levelt, C. C. (2016). Preliminary Experiments on Human Sensitivity to Rhythmic Structure in a Grammar with Recursive Self-Similarity. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00281

Geambaşu, A., Toron, L., Ravignani, A., & Levelt, C. C. (2020). Rhythmic Recursion? Human Sensitivity to a Lindenmayer Grammar with Self-similar Structure in a Musical Task. *Music & Science*, *3*, 205920432094661. https://doi.org/10.1177/2059204320946615

Goschke, T. (1998). Implicit learning of perceptual and motor sequences : Evidence for independent learning systems. In *Handbook of implicit learning* (p. 401-444). Sage Publications, Inc.

Hommel, B. (1994). Spontaneous decay of response-code activation. *Psychological Research*, *56*(4), 261-268. https://doi.org/10.1007/BF00419656

Huang, J., Dai, H., Ye, J., Zhu, C., Li, Y., & Liu, D. (2017). Impact of Response Stimulus Interval on Transfer of Non-local Dependent Rules in Implicit Learning : An ERP Investigation. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.02107

Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience*, *5*(5), Article 5. https://doi.org/10.1038/nn841

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*, 188–196. https://doi.org/10.1037/h0056940

Isbilen, E. S., & Christiansen, M. H. (2022). Statistical Learning of Language : A Meta-Analysis Into 25 Years of Research. *Cognitive Science*, *46*(9), e13198. https://doi.org/10.1111/cogs.13198

Krivochen, D., Phillips, B., & Saddy, J. (2018). *Classifying points in Lindenmayer systems : Transition probabilities and structure reconstruction (v. 1.1)*. https://doi.org/10.13140/RG.2.2.25719.88484

Kuhn, G., & Dienes, Z. (2006). Differences in the types of musical regularity learnt in incidental- and intentional-learning conditions. *Quarterly Journal of Experimental Psychology*, *59*(10), 1725-1744. https://doi.org/10.1080/17470210500438361

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1545. https://doi.org/10.1037/a0017010

Lindenmayer, A. (1968). Mathematical models for cellular interactions in development II. Simple and branching filaments with two-sided inputs. *Journal of Theoretical Biology*, *18*(3), 300-315. https://doi.org/10.1016/0022-5193(68)90080-5

Lynn, C. W., Kahn, A. E., Nyema, N., & Bassett, D. S. (2020). Abstract representations of events arise from mental errors in learning and memory. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-15146-7

McCarthy, G., & Donchin, E. (1981). A Metric for Thought: A Comparison of P300 Latency and Reaction Time. *Science*, *211*(4477), 77–80. https://doi.org/10.1126/science.7444452

Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical Implications of Articulatory Duration, Phonological Similarity, and Phonological Complexity in Verbal

Working Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1353-1380. https://doi.org/10.1037/0278-7393.29.6.1353

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

Norman, E., Price, M. C., & Duff, S. C. (2006). Fringe consciousness in sequence learning: The influence of individual differences. *Consciousness and Cognition*, *15*(4), 723–760. https://doi.org/10.1016/j.concog.2005.06.003

Norman, E., Price, M. C., Duff, S. C., & Mentzoni, R. A. (2007). Gradations of awareness in a modified sequence learning task. *Consciousness and Cognition*, *16*(4), 809-837. https://doi.org/10.1016/j.concog.2007.02.004

Oberauer, K. (2019). Working memory and attention—A conceptual analysis and review. *Journal of Cognition*, *2*. https://doi.org/10.5334/joc.58

Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. *Journal of Experimental Psychology: General*, *142*(2), 380–411. https://doi.org/10.1037/a0029588

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language*, *73*, 15-30. https://doi.org/10.1016/j.jml.2014.02.003

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2 : Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Perruchet, P., & Amorim, M.-A. (1992). Conscious knowledge and changes in performance in sequence learning : Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 785-800. https://doi.org/10.1037/0278-7393.18.4.785

Perruchet, P., Bigand, E., & Benoit-Gonin, F. (1997). The emergence of explicit knowledge during the early phase of learning in sequential reaction time tasks. *Psychological Research*, *60*(1), 4-13. https://doi.org/10.1007/BF00419676

Planton, S., Kerkoerle, T. van, Abbih, L., Maheu, M., Meyniel, F., Sigman, M., Wang, L., Figueira, S., Romano, S., & Dehaene, S. (2021). A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans. *PLOS Computational Biology*, *17*(1), e1008598. https://doi.org/10.1371/journal.pcbi.1008598

Pothos, E. (2010). An entropy model for artificial grammar learning. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00016

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/. (s. d.).

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radulescu, S., Kotsolakou, A., Wijnen, F., Avrutin, S., & Grama, I. (2021). Fast but Not Furious. When Sped Up Bit Rate of Information Drives Rule Induction. *Frontiers in Psychology*, *12*, 4987. https://doi.org/10.3389/fpsyg.2021.661785

Radulescu, S., Wijnen, F., & Avrutin, S. (2019). Patterns Bit by Bit. An Entropy Model for Rule Induction. *Language Learning and Development*, 1–32. https://doi.org/10.1080/15475441.2019.1695620

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests : Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 585-594. https://doi.org/10.1037/0278-7393.20.3.585

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay Theory of Immediate Memory: From Brown (1958) to Today (2014). *Quarterly Journal of Experimental Psychology*, *69*(10), 1969–1995. https://doi.org/10.1080/17470218.2014.914546

Savalia, T., Shukla, A., & Bapi, R. S. (2016). A Unified Theoretical Framework for Cognitive Sequencing. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.01821

Schmid, S., Saddy, D., & Franck, J. (2023). Finding Hierarchical Structure in Binary Sequences: Evidence from Lindenmayer Grammar Learning. *Cognitive Science*, *47*(1), e13242. https://doi.org/10.1111/cogs.13242

Schwarb, H., & Schumacher, E. H. (2012). Generalized lessons about sequence learning from the study of the serial reaction time task. *Advances in Cognitive Psychology*, *8*(2), 165-178. https://doi.org/10.5709/acp-0113-1

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1435-1451. https://doi.org/10.1037/0278-7393.25.6.1435

Shanks, D. R., Wilkinson, L., & Channon, S. (2003). Relationship between priming and recognition in deterministic and probabilistic sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 248-261. https://doi.org/10.1037/0278-7393.29.2.248

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shirley, E. J. (2014). *Representing and remembering Lindenmayer-grammars* [Ph.D., University of Reading]. https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.658878

Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance, 11*(5), 598. https://doi.org/10.1037/0096-1523.11.5.598

Soetens, E., Melis, A., & Notebaert, W. (2004). Sequence learning and sequential effects. *Psychological Research*, *69*(1), 124-137. https://doi.org/10.1007/s00426-003-0163-4

Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 674-685. https://doi.org/10.1037/0278-7393.21.3.674

Sternberg, S. (1969). Memory-Scanning: Mental Processes Revealed by Reaction-Time Experiments. *American Scientist*, *57*(4), 421–457. https://www.jstor.org/stable/27828738

Vender, M., Krivochen, D. G., Compostella, A., Phillips, B., Delfitto, D., & Saddy, D. (2020). Disentangling sequential from hierarchical learning in Artificial Grammar Learning : Evidence from a modified Simon Task. *PLOS ONE*, *15*(5), e0232687. https://doi.org/10.1371/journal.pone.0232687

Vender, M., Krivochen, D. G., Phillips, B., Saddy, D., & Delfitto, D. (2019). Implicit Learning, Bilingualism, and Dyslexia : Insights From a Study Assessing AGL With a Modified Simon Task. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01647

Verwey, W. B., & Wright, D. L. (2014). Learning a keying sequence you never executed : Evidence for independent associative and motor chunk learning. *Acta Psychologica*, *151*, 24-31. https://doi.org/10.1016/j.actpsy.2014.05.017

Verwey, W., & Dronkers, W. J. (2019). Skill in discrete keying sequences is execution rate specific. *Psychological research*. https://doi.org/10.1007/s00426-017-0967-2

Vitányi, P. M. B., & Walker, A. (1978). Stable string languages of lindenmayer systems. *Information and Control*, *37*(2), 134-149. https://doi.org/10.1016/S0019-9958(78)90483-7

Wilkinson, L., & Shanks, D. R. (2004). Intentional Control and Implicit Sequence Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 354-369. https://doi.org/10.1037/0278-7393.30.2.354

Willingham, D. B., Greenberg, A. R., & Thomas, R. C. (1997). Response-to-stimulus interval does not affect implicit motor sequence learning, but does affect performance. *Memory & Cognition*, *25*(4), 534-542. https://doi.org/10.3758/BF03201128