



HAL
open science

Probability, Dependency, and Frequency Are Not All Equally Involved in Statistical Learning

Laura Lazartigues, Fabien Mathy, Frédéric Lavigne

► To cite this version:

Laura Lazartigues, Fabien Mathy, Frédéric Lavigne. Probability, Dependency, and Frequency Are Not All Equally Involved in Statistical Learning. *Experimental Psychology*, 2023, 69 (5), pp.241-252. <10.1027/1618-3169/a000561>. <hal-04016782>

HAL Id: hal-04016782

<https://hal.science/hal-04016782v1>

Submitted on 8 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Probability, Dependency and Frequency are not all Equally Involved in Statistical Learning

Laura Lazartigues, Fabien Mathy, and Frédéric Lavigne

Department of Psychology, Université Côte d'Azur, BCL, CNRS, Nice, France

Author Note

Laura Lazartigues  <https://orcid.org/0000-0002-5361-647X>

Fabien Mathy  <https://orcid.org/0000-0002-7705-5689>

Frédéric Lavigne  <https://orcid.org/0000-0003-4493-430X>

The authors declare that they have no conflict of interest to disclose. Correspondence concerning this article should be addressed to Laura Lazartigues, Université Côte d'Azur, BCL UMR 7320, CNRS, Campus Saint Jean d'Angely, MSHS, 24 avenue des diables bleus, 06357 Nice Cedex 4, France. Email: Laura.lazartigues@univ-cotedazur.fr.

Abstract

The ability to learn sequences depends on different factors governing sequence structure, such as transitional probability (TP, probability of a stimulus given a previous stimulus), adjacent or nonadjacent dependency, and frequency. Current evidence indicates that adjacent and nonadjacent pairs are not equally learnable; the same applies to second-order and first-order TPs and to the frequency of the sequences. However, the relative importance of these factors and interactive effects on learning remain poorly understood. The first experiment tested the effects of TPs and dependency separately on the learning of nonlinguistic visual sequences and the second experiment used the factors of the first experiment and added a frequency factor to test their interactive effects with verbal sequences of stimuli (pseudo-words). The results of both experiments showed higher performance during online learning for first-order TPs in adjacent pairs. Moreover, Experiment 2 indicated poorer performance during offline recall for nonadjacent dependencies and low-frequency sequences. We discuss the results that different factors are not used equally in prediction and memorization.

Keywords: Statistical learning, Prediction, Sequence learning, Artificial words

Probability, Dependency and Frequency are not all Equally Involved in Statistical Learning

Adaptation to the environment depends on the ability to learn associations between stimuli in sequences (De Houwer et al., 2013). Statistical learning in particular has been shown to be a central process to learn such serial associations (Destrebecqz & Cleeremans, 2001) in different sensory modalities, whether auditory (Saffran, Newport, et al., 1996), visual (Fiser & Aslin, 2002) or tactile (Conway & Christiansen, 2005). Some abilities based on statistical learning appear as early as a few months old in humans, for instance in word segmentation (Saffran, Aslin, et al., 1996; Saffran & Thiessen, 2003), for the acquisition of language (Pelucchi et al., 2009b), mental grammar (Nemeth et al., 2011) and when learning abstract sequences (Kirkham et al., 2002).

Statistical learning of associations is reported to depend not only on the frequency of the elements of a learning context (Aslin, 2017), but also on the transitional probability between elements that can be adjacent (Saffran, Newport, et al., 1996) or nonadjacent, referring to nonadjacent dependencies (Newport & Aslin, 2004) and long-distance dependencies (Phillips et al., 2005). One issue is that frequency and transitional probability use different types of information in the sequence of stimuli, hence raising the question of their relative importance in learning. Transitional probability (TP) describes the conditional relation between the elements embedded in a set of sequences. For instance, in a sequence ABC , the transitional probability $p(B|A)$ of B given A corresponds to the number of occurrences of B following all occurrences of A in sequences containing both A and B . In other words, TP refers to the strength with which B can be predicted by A (Pelucchi et al., 2009a). TP has been shown to be more informative in predicting a stimulus given a previous stimulus than the frequency of the pair (Mirman et al., 2010). Furthermore, transitional

probability between nonadjacent elements $p(C|A)$ corresponds to the strength with which C can be predicted by A in a sequence ABC . This type of nonadjacent dependency has been reported to be more difficult to learn than adjacent dependency (Wilson et al., 2018).

Behavioral data has shown that nonadjacent dependencies cannot be learned in a divided attention task, while adjacent dependencies can (Pacton et al., 2015). A study using event-related potentials has highlighted that adjacent and nonadjacent dependencies are not equally learnable because nonadjacent dependencies involve a greater processing cost (Phillips et al., 2005). This higher processing cost is explained by Phillips et al. (2005) by a higher storage cost for nonadjacent dependencies. Thus, TP has been considered a fundamental process in statistical learning (Aslin, 2017), and it has been used extensively to determine how adjacent and nonadjacent elements are processed in diverse domains (e.g., artificial grammar learning with syllables, Gomez, 2002; Perruchet & Poulin-Charronnat, 2012; ASRT tasks, Janacsek et al., 2012). Statistical learning and transitional probability eventually emerged as crucial processes to account for how individuals acquire language (Romberg & Saffran, 2010; Thompson & Newport, 2007), how learned associations related to transitional probability allow for semantic predictions involved in priming (Brunel & Lavigne, 2009), how reading operates (Lavigne et al., 2000).

Although transitional probability appears central to language processing, but a major limitation in the literature is that studies have usually focused on TPs between pairs of stimuli. However, language involves statistical regularities outside the pairs (Lavigne et al., 2016). For instance, in a triplet of events such as ABC , C can be predicted by each of the preceding stimuli A and B but also by the combination of them (AB). This latter case involves the second-order transitional probability of C given the combination of A and B noted $p(C|AB)$ (Lavigne et al., 2014; also called second-order dependencies or conditionals (SOCs), see Gomez, 1997). Despite these previous studies, the effects of second-order TP on learning are

still poorly understood, particularly because they have been studied independently of the effect of other factors. Our hypothesis is that learning first-order vs. second-order TP could be better disentangled using sequences of three stimuli obeying the logic of the exclusive-OR (XOR, Lavigne et al., 2016).

The XOR is a logical rule that has initially been used in categorization studies (J. D. Smith et al., 2011). To illustrate its structure, let's imagine a set of four objects that differ according to two dimensions: shape (square or circle) and color (blue or yellow). The set of possible objects combining these two dimensions is: blue square, yellow square, blue circle, and yellow circle. By applying the logic of the XOR to these four objects, the categories could be structured this way: the first category contains square OR blue objects, BUT not blue squares (that is, category 1: yellow squares and blue circles; category 2: blue squares and yellow circles). The categories are therefore based on a relationship between the shape and color dimensions. Category 1 requires the presence of one of two diagnostic features but not both. This connective is called the eXclusive OR because it eXcludes the conjunction of the diagnostic features. The XOR would fit the “cheese OR dessert” options rather well in a french menu as the client is only allowed to pick one of two options (the two allowed options are those of category 1; the client cannot choose both options, and we imply that because the client paid for at least one choice, not picking one option is a second impossibility).

In a serial version of the XOR, the rule allows the creation of sequential triplets with the two input dimensions corresponding to the first stimulus and the second stimulus while the category corresponds to the third stimulus (fully predictable based on the two first stimuli). The serial organization of the dimensions makes the relational aspect of the XOR more straightforward for sequences of variables $A-B-C$. One example would be four sequences of two successive bits $ab, ab', a'b, a'b'$ predicting the third bit c vs c' . One XOR could be that the third stimulus C is c if and only if the two previous stimuli AB are either $a'b$

or ab' . The verbal rule would be: “Stimulus C is c if and only if Stimulus A is a XOR Stimulus B is b ; else c' ”, or in more intelligible terms “Stimulus C is c if and only if Stimulus A is a OR Stimulus B is b , BUT not both a and b ”. In terms of serial processing, one thus need to keep in mind the value of A when the value of B appears in order to combine both values to predict C .

Applying the XOR to triplets of serial stimuli ABC , we can notice that all first-order TPs (C given A and C given B) are less than one (neither A or B alone allow full prediction of the third stimulus C). However, the second-order TP (C given the combination of A and B) does allow exact prediction of C . The XOR logic makes second-order TP informative for prediction (not first-order TP) because all first-order TPs are equal to .5 whereas the second-order TP is equal to 1 (Rey et al., 2022). In other words, one cannot predict the third stimulus without having information on both A and B . We thus aimed to test the learning of second-order TPs and compare it to the learning of first-order TPs for adjacent or nonadjacent items in two experiments.

Moreover, because the precise learning differences between first- and second-order TP and between adjacent and nonadjacent dependencies remains poorly understood, the first experiment aimed to study the specific effect of these factors during statistical learning. Indeed, some studies have indicated that first-order TPs are less complex to learn than second-order TPs (Gomez, 1997), and that adjacent dependencies are easier to learn than nonadjacent dependencies (Pacton et al., 2015). Nevertheless, the use of different protocols has also suggested that sequences including second-order TPs and nonadjacent dependencies can be easily learned (Barnes et al., 2008; Tóth-Fáber et al., 2021). However, in all these previous studies, the factors of interest have not been all tested with the same material.

To better understand the specific effect of TP and dependency on learning, we used four different rules (described in more detail in the method section), each governing four

sequences of nonlinguistic visuospatial stimuli. We predicted that learning should be more efficient for the conditions involving first-order TPs in adjacent dependencies, followed by the condition involving second-order TPs (because, despite the higher complexity, the dependency between the combination *AB* and the item *C* remained adjacent) and then poorer performance for the condition involving nonadjacent dependencies.

To go further, our second experiment aimed to study the same factors (first-order TPs for adjacent or nonadjacent dependencies and second-order TPs) in interaction in a more complex task that used sequences of language-related stimuli made of syllables. The frequency of the sequences was also manipulated, given that frequency has also been reported to have a role during statistical learning (Thiessen et al., 2013). We hypothesized that the effects of TPs, adjacency and frequency would all influence the learning of language-related stimuli despite the higher complexity of the task. We further hypothesized that even though learning could take longer to occur in this complex design, the three factors could all contribute to learning and that they would interact. We posited that the effects should be the same as in Experiment 1 with a precedence for first-order TPs in adjacent dependencies, and we hypothesized better performance with higher frequency.

Experiment 1

Method

Participants. Eighty students between 18 and 30 years old (mean age 22.4) participated in the experiment, voluntarily or in exchange for course credits. Each condition was completed by twenty participants. The experiment was approved by the local ethics committee (CERNI) of Université Côte d'Azur, and the experiment was conducted with the informed written consent of the participants.

Material. In the present study, we used sequences of three red dots (with each

stimulus characterized by its location on a screen) of the form $A - B - C$, with A , B and C being presented in this temporal order. This allowed us to embed both first-order TPs between adjacent pairs ($p(B|A)$ and $p(C|B)$) as well as between nonadjacent pairs ($p(C|A)$). This further allowed us to embed second-order TPs ($p(C|AB)$) between the triplets of stimuli.

Four conditions were created, and each participant completed only one of the four conditions. In the first condition, four triplets were used in which the first item predicted the second item with certainty ($p(B|A) = 1$) and two items were possible for position C ($p(C|B) = 0.5$). This condition was called AB^* (event C is replaced by an asterisk since it is not predictable). This condition allowed us to test the learning of first-order TPs between the two adjacent elements of the first pair. In the second condition (called $*BC$), the second item allowed us to predict the last item ($p(C|B) = 1$), but the first item did not allow the prediction of the second item ($p(B|A) = 0.5$). The third condition (called A^*C) was thought to test the learning of first-order TPs between two nonadjacent items, with the first item of the sequence predicting the last item with certainty ($p(C|A)=1$). However, a random second item produced $p(B|A) = 0.5$ and $p(C|B) = 0.5$ by construction. The fourth condition consisted of four sequences combined according to an exclusive-OR (XOR) logic (Lavigne et al., 2014). The XOR allowed us to dissociate first-order TPs ($p(B|A)$, $p(C|B)$, and $p(A|C)$) and second-order TPs ($p(C|AB)$). While the third item could not be fully predicted by either of the two preceding items considered separately, the third item could be fully predicted by the combination of the first two items. The XOR hence allowed us to study the learning of second-order TPs by maintaining a constant first-order TP.

The experiment was programmed in PsychoPy (Peirce, 2007), and stimuli were displayed on a touchscreen. Nine positions were marked by nine crosses '+' displayed on a 3 X 3 virtual grid on a gray background. Each sequence was made of three red dot stimuli that replaced the crosses successively at different positions when the participant touched the

stimuli. A preliminary phase was carried out to select the most neutral sequences that could be used during the experiment. In this pretest, 10 separate participants completed 800 trials of random triplets. Transition times between two successive touches were recorded. To constitute the triplets to be used in Experiment 1, we selected four pairs of stimuli for each transition (the first transition between the first two items and the second transition between the last two items), allowing us to have only two items in the first position, two others in the second position and again two others in the third position. Moreover, the item at the end of the first transition and at the beginning of the second had to be the same (e.g. 9-3 and 3-7) with similar response times. Then, triplets were created with the conjunction of these pairs in accordance with the condition (Table 1).

Protocol and task. We used the same task as Rey et al., (2019) but with sequences of three items. A trial started when the participant touched a yellow cross always displayed at the bottom of the screen. The yellow cross then disappeared and was replaced by the first red dot displayed at one of the nine possible locations. When the participant touched the first red dot, it was instantly replaced by the second red dot at one of the eight remaining locations (i.e., the inter-stimulus interval (ISI) between dot stimuli was set to zero milliseconds). The second red dot once touched was then instantly replaced by the third red dot at one of the seven remaining locations.

Participants were instructed to touch the red dots as quickly as possible. Whenever the participant touched the second or the third red dot within a sequence, the response time (RT) was recorded as TT1 (transition time 1, time to touch the second dot after touching the first dot) and TT2 (transition time 2, time to touch the third dot after touching the second dot). The next trial began immediately after the completion of a sequence by touching the third red dot, with the yellow cross being displayed again at the bottom of the screen.

Each task was made of 400 trials divided into 10 blocks of 40 trials (10 trials for each

sequence by block). Each trial consisted of presenting one of the four sequences of the chosen condition for a given participant. At the end of the 400 trials of the learning phase, a switch phase occurred. The switch phase consisted of the permutation of the predictable stimulus (the second red dot for the AB^* condition and the third red dots for other conditions) between the different sequences (Table 1). The switch phase allowed us to check whether the sequences had been learned. A decrease in RTs for the predictable transitions ($p = 1$) was expected during the learning phase, but it is only an increase in RTs during the switch phase that allowed us to conclude that a learning process took place instead of a simple training for the task (that is, participants becoming faster at touching the selected dots in their condition).

Results and discussion

Data are posted at <https://osf.io/hjy9k/>. We removed 2% of the data corresponding to response times (RT) faster than 150 ms and slower than 850 ms (± 2 standard deviations from the mean).

Global analyses. Figure 1 shows a decrease in RTs during the learning phase (first 10 blocks) for all transitions of each condition but with a stronger decrease for the predictable transitions ($p = 1$) based on first-order TPs for adjacent pairs (AB^* and $*BC$). An increase in RTs also appeared between the last block of the learning phase (Block 10) and the switch phase (last block) for all predictable transitions, indicating effective learning of the structure rather than a simple automatization of responses due to a training in the task.

Linear mixed models (lmm) with random intercepts by participant were carried out on the RTs within the learning phase (first 10 blocks) using the lme4 package for R (Team, 2013). The factors blocks (1 to 10), condition (AB^* , $*BC$, A^*C and XOR) and predictability (the present item is predictable given past items vs. not) were tested. The results in Table 2 indicate a significant decrease in RTs during the learning phase but with a difference of steepness as a function of the condition and the predictability of the transition. These findings

suggest a different effect based on the type of TP (first-order vs. second-order, $d = .21$) or dependency (adjacent vs. nonadjacent, $d = .18$) with better performance for sequences involving first-order TP in adjacent dependencies. Moreover, a Bayesian ANOVA suggests that the model that included all factors (Condition + Prediction + Block + Condition * Prediction) showed evidence against the null hypothesis ($BF_{10} = 3.370e+166$). For the switch phase, a lmm with the factors phase (last block of learning phase vs. switch block) and condition indicated an increase in RTs for the predictable transitions ($F(1,114) = 67.206$, $p < .001$, $d = 1.37$) and a significant interaction with the condition, showing a greater increase of RTs with the switch phase for the conditions AB^* and $*BC$ compared to the conditions A^*C and XOR ($F(3,114) = 12.836$, $p < .001$, $d = -.82$). We note that no effect of the switch phase appeared on the unpredictable transitions ($F(1,108) = 0.107$, $p = .744$, $d = .01$). Bayesian ANOVAs indicated that the best model included the factors condition, block and their interaction for predictable transitions ($BF_{10} = 1.536e+15$ against null hypothesis) but only the factor condition for unpredictable transitions ($BF_{10} = 2.876$). These results correspond to our expectations because the sequences used in the switch phase only involved a modification of items in the predictive transition (Table 1), so the items in unpredictable transitions remained the same as those in the learning phase.

Analyses by condition. To better understand the effect of TPs and adjacency, we focused on our four conditions. The first condition of interest was the AB^* condition, corresponding to a set of sequences where the first stimulus allowed us to predict the second stimulus (Figure 2a), and we specifically studied the evolution of RTs of TT1 (measured at the second stimulus) for this condition. In this condition, there was only one piece of information for participants to take into account: the probability of B knowing A . The second condition was more complex; indeed, in the $*BC$ condition, more elements were present: stimulus C was predicted by B but also by A in a nonadjacent way in sequence ABC

(Figure 2b). Thus, the **BC* condition should have faster RTs for its TT2 than the TT1 of the *AB** condition because the stimulus to be predicted received more activation (Lavigne et al., 2011). A first analysis including block and condition factors indicated a significant effect of the condition when only *AB** and **BC* were compared during the learning phase with shorter RTs for the **BC* condition ($F(1,377) = 13.18, p < .001, d = .37$) but no significant effect of the interaction block*condition ($F(1,366) = 0.926, p = .336, d = -.1$). Moreover, Bayesian analysis indicated that the best model only included the block factor ($BF_{10} = 1.738e+39$ against null hypothesis, $BF_m = 33.961$). This result suggested a weak effect of the nonadjacent dependency for the prediction because the **BC* condition did not lead to shorter RTs through the learning phase than the *AB** condition. However, in the **BC* condition, the TP of *C* knowing *A* was equal to .5, so the next condition allowed us to test the effect of the nonadjacent TP when its value was higher.

The *A*C* condition had all adjacent TPs equal to .5 and the nonadjacent TP equal to 1 (Figure 2c) to test the effect of the nonadjacent dependency during prediction. An analysis indicated no effect of the transition during the learning phase for the *A*C* condition ($F(1,19) = 0.657, p = .428, d = -.37$) however a significant effect of the switch phase appeared for TT2 ($F(1,19) = 4.899, p = .039, d = 1.02$) but not for TT1 ($F(1,19) = 0.557, p = .465, d = .34$). These results indicated a weak learning of the nonadjacent dependency. Indeed, this TP did not allow a prediction leading to shorter TT2 than TT1 during the learning phase. Nevertheless, the effect of the switch phase indicated learning of this TP, but the pattern during the learning phase suggested difficult learning. This result also explains why the same pattern appeared for the conditions *AB** and **BC*: the nonadjacent dependency in the **BC* condition was not sufficient to provide an advantage compared to the *AB** condition.

The *XOR* condition allowed us to test the effect of the second-order TP. In this

condition, all first-order TPs were equal to .5, and only the combination of the first two stimuli allowed us to predict the last stimulus (Figure 2d). The results indicated the same pattern as the $A*C$ condition: no effect between the RTs of TT1 and TT2 during the learning phase ($F(1,19) = 0.142, p = .711, d = .17$) but a significant effect of the switch phase for TT2 ($F(1,19) = 6.736, p = .018, d = 1.19$) and not for TT1 ($F(1,19) = 0.231, p = .636, d = -.22$). These analyses indicated a weak learning of the second-order TP.

To better understand the respective effects of these factors in a more realistic learning context, it was decided to create a more complex task. For this purpose, it was chosen to group all our conditions in the same protocol, to test linguistic stimuli allowing more information to be measured (i.e., accuracy in addition to RTs). We also decided to add the frequency as a factor in Experiment 2 because it could provide more information on factors implied in statistical learning.

Experiment 2

Experiment 1 suggested that all types of TPs and adjacency can be learned by participants with an advantage for first-order TPs in adjacent dependencies. However, Experiment 1 tested only one condition for each participant, thus manipulating the factors separately as between-participants variables. Experiment 2 investigated the learning of these conditions mixed together as within-participant variables. This was made easier by using pseudo-words instead of a matrix of dots. Indeed, a matrix of dots would have embedded 24 dots and would have become too large and thus too confusing, while different pseudo-words could be easily constructed and memorized. Each pseudo-word was constructed according to one of the experimental conditions, and all were presented in the tasks. Only a few studies have addressed the combined effect of both TPs and frequency to test for their interactive effects (Endress & Langus, 2017; Lazartigues et al., 2021; Mirman et al., 2010; Perruchet & Poulin-

Charronnat, 2012) or by fixing the frequency parameter to show an effect of TPs (Aslin et al., 1998). These studies have suggested a prevalence of TPs over frequency in both prediction and memorization. Additionally, the three factors (i.e., TP, dependency and frequency) have generally been studied separately in the statistical learning of sequences. However, the fact that TP and frequency are learned on the basis of the same set of sequences suggests that they could interact, which would be a new interesting result to understand the benefits of each factor for the task involved (prediction or memorization).

In order to test the interaction between TPs and frequency, we decided to replace the AB^* condition. In Experiment 1, conditions AB^* and $*BC$ focused on the same factor (i.e., first-order TP for an adjacent pair) and the results showed that they produced similar performance, so we chose to drop the AB^* condition in Experiment 2. To replace it, we chose a condition involving an XOR with unbalanced sequences. This type of condition is interesting because it allowed us to disentangle the effects of frequency of the sequences and first-order TP. This was made possible by manipulating the frequency of sequences for fixed values of second-order TP (Lazartigues et al., 2021).

In this experiment, two different tasks were designed to focus on the prediction and memorization of the sequence.

We hypothesized, as in Experiment 1, that the conditions with first-order TPs in adjacent dependencies should produce better performance, followed by the condition involving second-order TPs with a manipulation of frequency (because an unbalanced division of frequencies is known to allow better learning, L. B. Smith et al., 2018), followed by the second-order TP without manipulation of frequency and finally by the condition with nonadjacent dependencies. Moreover, according to the literature (Endress & Langus, 2017; Lazartigues et al., 2021; Mirman et al., 2010), we hypothesized the prevalence of TPs over frequency during the learning.

Method

Participants. Thirty Université Côte D'Azur psychology students between 18 and 32 years of age (mean age 21.9) participated in the experiment voluntarily (20 among them who could not receive course credits obtained 50€ for their participation). In the present study, 33 students completed the experiment, but three of them were excluded due to a data logging problem (for one participant) or interruption during sessions resulting in biased performance. The experiment was approved by the local ethics committee (CERNI) of Université Côte d'Azur, and the experiment was conducted with the informed written consent of the participants.

Material and task. Participants were exposed to 16 pseudo-words made of three simple meaningless syllables (consonant-vowel, see Table 3). These 16 pseudo-words corresponded to four conditions governed by rules (described later) allowing us to test the effect of transitional probabilities (first-order and second-order), contiguity (adjacent dependency and nonadjacent dependency) and frequency of the triplet. For all pseudo-words, the last syllable was fully predictable based on TPs. The **BC*, *A*C* and *XOR* conditions were constructed with the same rules as in Experiment 1. The new condition (called *XOR Unbalanced*) allowed us to test the effect of second-order TPs (typical of the XOR) while including a manipulation of frequency: the first two pseudo-words were presented five times per block (the same as the pseudo-words in other conditions), the third one was presented twice per block and the last one was presented eight times per block. This corresponded to frequencies of 0.0625 for two pseudo-words, 0.025 for the least frequent pseudo-word and 0.1 for the most frequent pseudo-word.

In each condition, only six syllables were used to create pseudo-words. Different syllables were used in the different conditions, for a total of 24 syllables used in the

experiment. Ten different sets of pseudo-words were used and assigned randomly to the different participants. Note that pseudo-words were constructed so that only two vowels appeared in the last syllable, allowing a detection task on the third element (fully predictable).

Participants had to perform a target detection task on the last vowel appearing in the last syllable of each pseudo-word. This last vowel could be, for example, either “e” or “u”, each one being assigned a specific key. Participants had to press the “q” key if they saw the “e” vowel and press the “m” key if they saw the “u” vowel (note that the target vowels were not the same for all participants). Participants had to press the key as fast as possible when the third vowel of the sequence appeared. Correct answer rates and response times were recorded. All pseudo-words were presented in a random order in each block, and each occurred five times per block (except for two pseudo-words of the *XOR Unbalanced* condition presented twice or eight times per block to manipulate the frequency of the triplet of syllables). A block thus involved 80 trials. Each trial began with a white cross in the center of the screen, followed by the sequence of three syllables. To avoid an effect of a fixed rhythm of presentation, two parameters of the presentation of the stimuli were manipulated: the position of each syllable on the screen (half of the syllables had the consonant in the center, the other half had the vowel in the center) and the display time of the syllable (the presentation time of each syllable randomly varied between 250 ms and 400 ms with a fixed blank screen of 250 ms between each syllable). Each participant completed five sessions of 10 blocks, one session per day for five days. A switch phase of 80 trials was added at the end of the last session (corresponding to a switch of the last stimuli as in Experiment 1). After the experiment, to further check for actual learning, a two-alternative forced-choice (2AFC) task was conducted that requested participants to choose between the learned pseudo-words and lures constructed by using the same syllables as in the learned pseudo-words but in different combinations.

Results and discussion

Accuracy and RTs (task one) and recognition rates (task two) were analyzed by fitting linear mixed-effect models to the data with random intercepts by participant. We used the lmer function in the lme4 package for R (Team, 2013). To analyze RTs, we selected correct answers only.

A lmm only including the session factor showed a significant increase in accuracy ($F(1,119) = 21.534, p < .001, d = .85$) and a significant decrease in RTs ($F(1,119) = 48.513, p < .001, d = -1.28$) during the learning phase (i.e., the first five sessions without the switch phase). This improvement during the learning phase was followed by a general drop in performance between the last block of the learning phase and the switch phase (correct answer rates: $F(1,29) = 10.634, p < .003, d = -1.21$; RTs: $F(1,29) = 9.094, p < .005, d = 1.12$). Bayesian ANOVAs indicated that the model including the session factor showed evidence against the null hypothesis both for learning phase (accuracy : $BF_{10} = 1323$; RTs : $BF_{10} = 2.877e+6$) and switch effect (accuracy : $BF_{10} = 17.7$; RTs : $BF_{10} = 7.3$).

Figures 3a and 3b show an increase in performance during the learning phase but small differences between conditions. The results of a lmm including the factors session (without switch phase) and condition indicated a significant effect of the sessions with an increase in the correct answer rate during learning phase ($F(1,563) = 64.068, p < .001, d = .41$) but no effect of the condition ($F(3,563) = 1.938, p = .122, d > .10$ and $< .18$). However, when we compared the three conditions A^*C , XOR and XOR *Unbalanced* mixed together against $*BC$ alone, the effect fell short of significance $p = .056$ ($d = -.23$) with the $*BC$ condition showing higher correct answer rates across learning phases. A lmm again including session and conditions indicated a significant effect of the session on RTs ($F(1,563) = 159.155, p < .001, d = -.57, BF = 1.13$) with a decrease in RTs across the sessions ($estimates = -2.625e-02$) but no effect of the condition or interaction ($F(3,563) = 0.244,$

$p = .865$, $d < .1$, and $F(3,563) = 0.634$, $p = .593$, $d < .1$). Nevertheless, Bayesian ANOVAs indicated that the model including condition and session showed evidence against the null hypothesis (accuracy : $BF_{10} = 4.419e+22$, $BF_m = 17.12$; RTs : $BF_{10} = 4.895e+29$, $BF_m = 7.35$).

To study the effect of frequency and to compare it to that of TPs, we selected the *XOR Unbalanced* condition because this condition allows a partial dissociation of frequency and first-order TPs. Figures 3c and 3d show an increase of performance during the learning phase for correct answer rates and RTs, with poorer performance for the pseudo-words with the lowest first-order TPs and frequency. To determine which factors or combination of factors was primarily involved in sequential learning, we first ran three mixed models, all including the effect of sessions with one or more of the factors of interest (i.e., the first two models involved only frequency or first-order TP, and the last included both of them) on correct answer rates. Then, we computed the Akaike information criterion (AIC, Akaike, 1987) for each of them to determine which model best reflected the data. The results indicated that the best model included only first-order TP (smallest AIC, indicating the most parsimonious model; $AIC = 1098$ vs. $AIC = 1113$ for the two other models). The results indicated a significant effect of the session and an interaction between session and first-order TP on answer rates ($F(1,563) = 11.46$, $p < .001$, $d = -.13$ and $F(3,563) = 4.895$, $p = .002$, $d > .21$ and $< .30$). Regarding RTs, we only found a significant effect of the session ($F(1,563) = 131.46$, $p < .001$, $d = .43$). However, Bayesian analyses showed that the model including condition and session indicated evidence against the null hypothesis (accuracy : $BF_{10} = 17784$, $BF_m = 7.5$; RTs : $BF_{10} = 1.683e+27$, $BF_m = 11.82$). Moreover, when only sequences with equal frequency (i.e., 0.0625) were selected, the results showed an effect of TPs at the end of the learning phase ($F(1,29) = 4.11$, $p = .052$, $d = .75$). This finding reaffirms the crucial role of TPs and provides evidence that even if second-order TPs allowed an exact

prediction (such as in the sequences tested), the first-order TP between the last two items had an impact on learning and suggests a complex interaction of first-order and second-order TPs during the statistical learning of sequences with a second-order TP not fully learned when the first-order TP was not high enough.

Finally, Figures 3e and 3f show the performance of the 2AFC task presented at the end of the experiment, with poor performance for the A^*C condition and the weakest pseudo-words of the *XOR Unbalanced* condition. A lmm on all conditions indicated a significant effect of the condition on the recognition rate ($F(3,90) = 5.331, p = .002, d > .01$) with a lower recognition rate for the A^*C condition.

As before, the *XOR Unbalanced* condition was selected and three mixed models were tested on recognition rates: frequency only, first-order TP only, and the two factors combined.

The best model on this analysis only included frequency ($AIC = 179$ against 183 for the two other models), although this model did not show a significant effect for frequency ($F(2,91) = 2.728, p = .071, d > .37$ and $<.47$) and the inclusion of frequency factor did not increase the probability of the null model in the Bayesian analysis ($BF_m < 1$). Nevertheless, when we tested only the least frequent pseudo-words compared to the other three, an effect of frequency appeared short of significance ($F(1,30) = 3.824, p = .06, d = .71$), suggesting a lower recognition rate for the least frequent pseudo-words. Hence, although TP prevails over frequency in online prediction, visible in correct answer rates, frequency could be a more important factor for memorization, visible in recognition rates.

General discussion

The purpose of this study was to investigate the learning of (1) first-order TP in cases of adjacent or nonadjacent dependencies and (2) second-order TP as a function of sequence frequency and first-order TP. The two experiments used different protocols and materials to

investigate the relative importance of these factors as a function of visuospatial or language-related stimuli and in online prediction tasks or an offline recognition task.

In Experiment 1, first-order TPs in adjacent pairs exhibited stronger effects on RTs with results showing the pattern of performance ($AB^*, *BC$) > (A^*C, XOR), meaning that the conditions AB^* and $*BC$ superseded the two others. Indeed, predictable transitions ($p = 1$) for the AB^* and $*BC$ conditions led to shorter RTs than tasks in which contiguous pairs were made of unpredictable transitions ($p = .5$). As in previous studies, our results indicate more efficient learning of first-order TP compared to second-order TP (Gomez, 1997) and adjacent dependencies compared to nonadjacent dependencies (Pacton et al., 2015; Phillips et al., 2005). Indeed, our finding suggests that participants used predominantly first-order TPs between two adjacent elements during sequence processing to predict the next stimulus to appear in the sequence and, thus enabling them to be faster. Conversely, nonadjacent dependencies (A^*C condition) as well as second-order TPs (XOR condition) did not lead to faster RTs for the predictable transition during the learning phase, although a drop in performance appeared with sequence modification. These results indicate difficult learning for both nonadjacent dependencies and second-order TPs.

The results of Experiment 2 showed that pseudo-words in all conditions were learned by participants, with a stronger effect of first-order TPs in adjacent dependencies on correct answer rates with ($*BC$) > ($A^*C, XOR, XOR Unbalanced$). However, the results differed for the recognition task, with only poor performance for nonadjacent dependencies such as ($*BC, XOR, XOR Unbalanced$) > (A^*C). Thus, nonadjacent pairs appeared to be less well recalled, even though the detection task indicated learning of these dependencies. The result that nonadjacent pairs led to poorer recognition performance highlights the difficulties of learning this type of dependency.

The fact that the results of both experiments indicate a prevalence of first-order TPs in

adjacent pairs points to the possibility of similar mechanisms involved in learning sequences of nonlinguistic visuospatial stimuli and sequences of language-related stimuli. Moreover, this concurs with the idea of domain-general learning mechanisms (Perruchet & Pacton, 2006).

Note that this does not mean that these two types of information are learned by the same brain structure or region, but that similar synaptic and neural mechanisms may be at work in learning sequences of the two types of stimuli.

Turning to the *XOR Unbalanced* condition in Experiment 2, the two tasks (detection task and 2AFC task) focusing either on online prediction or on memorization showed stronger effects of TPs or frequency, respectively. This suggests that both TP and frequency were learned during the experimental sessions but that participants did not use the same factor in both tasks. These results are in contrast with previous studies showing a prevalence of TP over frequency for both prediction (Mirman et al., 2010) and memorization (Endress & Langus, 2017). The difference between our results and those of Endress and Langus (2017) could be due to the type of stimulus presentation. Indeed, in our recognition task, the pseudo-words were presented visually as units (i.e., the components of the pseudo-words were not presented sequentially) whereas Endress et al. used an auditory serial presentation. This difference could explain our results. Since a non-sequential presentation does not require any prediction, the frequency could show an effect stronger than TP. In addition, low frequency led to poorer performance for recognition and this result is in line with the better recognition of frequent words than of rare words in reading (Albregues et al., 2019). Moreover, Experiment 2 brings new results on learning TP for linguistic stimuli compared to results from (Lazartigues et al., 2021) for nonlinguistic stimuli. The Lazartigues et al. (2021) study involved sequential pointing of dots on a touch screen, similar to Experiment 1 but with sequences obeying an XOR Unbalanced rule comparable to that of Experiment 2 (with the same values of TPs). Their results indicated a prevalence of TPs over frequency, as in

Experiment 2. However, in Lazartigues et al. (2021), the RTs between the second and third stimuli of the sequences did not show an effect of TP for high-frequency sequences, while Experiment 2 showed an effect of TP on correct answer rates even for high-frequency sequences.

The use of different protocols to study the same factors reaffirms the question of the differential role of TPs depending on the task. The results of Experiment 2 point to the fact that the information used by participants depends on the task at stake. It is interesting to note that for the same set of sequences that are learned, participants use in greater proportion either the TP between stimuli processed in the sequence or the frequency of the triplets. The present study highlights the relation between task properties and the information used. On the one hand, the online task of Experiment 2 allowed prediction of a stimulus from the preceding one(s). In that case, the TP was the information used by participants with a greater effect of first-order TP on performance and second-order TP appearing to be more difficult to learn even though it allowed perfect prediction. On the other hand, the 2AFC task of Experiment 2 required offline recognition of triplets of stimuli presented at the same time and hence did not require predicting a stimulus from the preceding stimuli. In that case, the frequency of the sequences had a stronger effect on recognition than TP. This new result that TPs are more used by participants in prediction and frequency in recognition suggests that the effects of TP and frequency do not depend on different learning stages but rather on different uses of the learned information. When prediction can help online processing of sequentially presented stimuli, participants use the TP, which is information about which stimulus will come next in the sequence. When recall is tested offline, the effect of frequency arises and influences recognition, as is the case in word reading. The task-dependent relative effects of TP and frequency on the basis of the same learned set of sequences are in line with Thiessen et al. (2013), who reported that frequency and TP are involved in different types of statistical

learning mechanisms. The results of the present study point to the possibility that, based on a single set of learned parameters, including TP and frequency, task-dependent mechanisms can lead to the use of TPs or frequency, depending on whether the task requires recognition or prediction.

References

- Akaike, H. (1987). Factor analysis and AIC. In *Selected papers of hirotugu akaike* (p. 371-386). Springer.
- Albregues, C., Lavigne, F., Aguilar, C., Castet, E., & Vitu, F. (2019). Linguistic processes do not beat visuo-motor constraints, but they modulate where the eyes move regardless of word boundaries : Evidence against top-down word-based eye-movement control during reading. *PloS one*, *14*(7), e0219666.
- Aslin, R. N. (2017). Statistical learning : A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1-2), e1373.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science*, *9*(4), 321-324.
- Barnes, K. A., Howard Jr, J. H., Howard, D. V., Gilotty, L., Kenworthy, L., Gaillard, W. D., & Vaidya, C. J. (2008). Intact implicit learning of spatial context and temporal sequences in childhood autism spectrum disorder. *Neuropsychology*, *22*(5), 563.
- Brunel, N., & Lavigne, F. (2009). Semantic priming in a cortical network model. *Journal of Cognitive Neuroscience*, *21*(12), 2300-2319.
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24.
- De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic bulletin & review*, *20*(4), 631-642.
- Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic bulletin & review*, *8*(2), 343-350.

- Endress, A. D., & Langus, A. (2017). Transitional probabilities count more than frequency, but might not be used for memorization. *Cognitive psychology*, *92*, 37-64.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.
- Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, *33*(2), 154-207.
- Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431-436.
- Janacek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills : Age-related differences in implicit sequence learning across the human lifespan. *Developmental science*, *15*(4), 496-505.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy : Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.
- Lavigne, F., Avnaïm, F., & Dumercy, L. (2014). Inter-synaptic learning of combination rules in a cortical network model. *Frontiers in Psychology*, *5*, 842.
- Lavigne, F., Dumercy, L., & Darmon, N. (2011). Determinants of multiple semantic priming : A meta-analysis and spike frequency adaptive model of a cortical network. *Journal of cognitive neuroscience*, *23*(6), 1447-1474.
- Lavigne, F., Longrée, D., Mayaffre, D., & Mellet, S. (2016). Semantic integration by pattern priming : Experiment and cortical network model. *Cognitive neurodynamics*, *10*(6), 513-533.
- Lavigne, F., Vitu, F., & d'Ydewalle, G. (2000). The influence of semantic context on initial eye landing sites in words. *Acta Psychologica*, *104*(2), 191-214.

- Lazartigues, L., Mathy, F., & Lavigne, F. (2021). Statistical learning of unbalanced exclusive- or temporal sequences in humans. *Plos one*, *16*(2), e0246826.
- Mirman, D., Graf Estes, K., & Magnuson, J. S. (2010). Computational modeling of statistical learning : Effects of transitional probability versus frequency and links to word learning. *Infancy*, *15*(5), 471-486.
- Nemeth, D., Janacsek, K., Csifcsak, G., Szvoboda, G., Howard Jr, J. H., & Howard, D. V. (2011). Interference between sentence processing and probabilistic implicit sequence learning. *PLoS One*, *6*(3), e17577.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive psychology*, *48*(2), 127-162.
- Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based associative account of adjacent and nonadjacent dependency learning valid? *Acta Psychologica*, *157*, 195-199.
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8-13.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse : Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244-247.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child development*, *80*(3), 674-685.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning : One phenomenon, two approaches. *Trends in cognitive sciences*, *10*(5), 233-238.
- Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations : Extracting word-like units when only statistical information is available. *Journal of Memory and Language*, *66*(4), 807-818.
- Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic

- long-distance dependencies. *Cognitive Brain Research*, 22(3), 407-428.
- Rey, A., Fagot, J., Mathy, F., Lazartigues, L., Tosatto, L., Bonafos, G., Freyermuth, J., & Lavigne, F. (2022). Learning Higher-Order Transitional Probabilities in Nonhuman Primates. *Cognitive Science*, 46(4), e13121.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906-914.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation : The role of distributional cues. *Journal of memory and language*, 35(4), 606-621.
- Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental psychology*, 39(3), 484.
- Smith, J. D., Coutinho, M. V., & Couchman, J. J. (2011). The learning of exclusive-or categories by monkeys (*Macaca mulatta*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, 37(1), 20.
- Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, 22(4), 325-336.
- Team, R. C. (2013). *R : A language and environment for statistical computing*.
- Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework : A two-process account of statistical learning. *Psychological bulletin*, 139(4), 792.
- Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax : The role of transitional probability. *Language learning and development*, 3(1), 1-42.
- Tóth-Fáber, E., Janacsek, K., Szöllösi, Á., Kéri, S., & Nemeth, D. (2021). Regularity detection under stress : Faster extraction of probability-based regularities. *Plos one*, 16(6),

e0253123.

Wilson, B., Spierings, M., Ravnani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., Van der Kant, A., Smith, K., & Rey, A. (2018). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*.

Table 1*Sequences used for the learning phase and switch phase for all conditions in Experiment 1.*

Condition	<i>AB</i> *	* <i>BC</i>	<i>A</i> * <i>C</i>	<i>XOR</i>
Learning phase	9-3-7	6-3-7	9-2-4	9-3-4
	9-3-4	9-3-7	9-3-4	9-2-7
	6-2-7	6-2-4	6-2-7	6-3-7
	6-2-4	9-2-4	6-3-7	6-2-4
Switch phase	9-2-7	6-3-4	9-2-7	9-3-7
	9-2-4	9-3-4	9-3-7	9-2-4
	6-3-7	6-2-7	6-2-4	6-3-4
	6-3-4	9-2-7	6-3-4	6-2-7

Note. The switch phase corresponds to a modification of the predictable item, i.e., the second for the *AB** condition and the last one for other conditions.

Table 2*Effects of block, condition and predictability on RTs during the learning phase in Experiment 1.*

Variable	Sum Sq.	Mean Sq.	NumDF	DenDF	F value	<i>p</i>
Block	851506	851506	1	1542.8	413.662	< .001***
Condition	56328	18776	3	1566.7	9.121	< .001***
Predictability	34430	34430	1	1542.8	16.726	< .001***
Block × Condition	29362	9787	3	1542.8	4.755	.003**
Block × Predictability	31438	31438	1	1542.8	15.273	< .001***
Condition × Predictability	20737	6912	3	1542.8	3.358	.018*
Block × Condition × Predictability	54826	18275	3	1542.8	8.878	< .001***

Note. The predictability factor corresponds to the possibility of predicting the present item from the previous items.

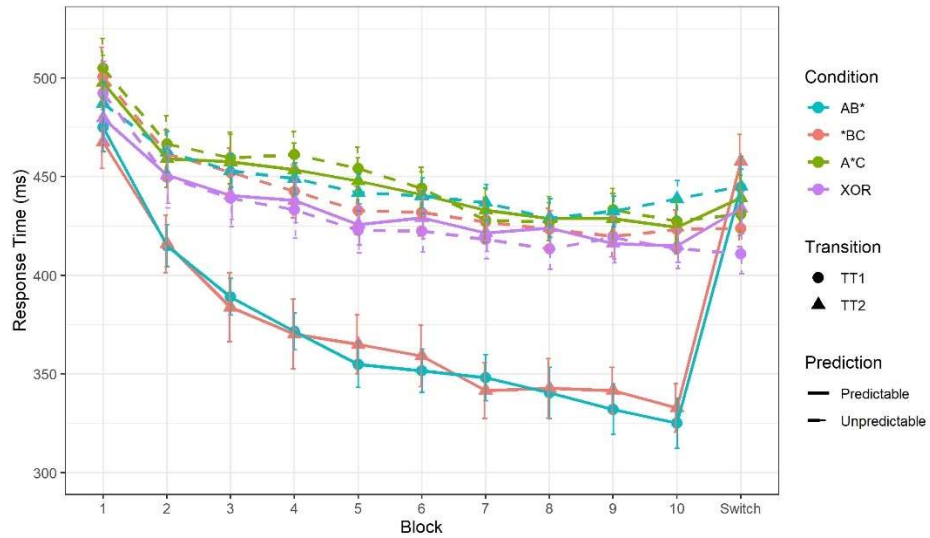
Table 3*Example of pseudo-words used for the learning phase and switch phase for all conditions in Experiment 2.*

Condition	* <i>BC</i>	<i>A</i> * <i>C</i>	<i>XOR</i>	<i>XOR Unbalanced</i>
Learning phase	PA-TI-DE	TA-MO-FU	DI-PO-PE	TO-KO-RE (frequency +)
	PA-RO-VU	TA-JI-FU	DI-GA-BU	TO-RA-LU (frequency =)
	MI-TI-DE	JO-MO-TE	FA-PO-BU	GO-KO-LU (frequency =)
	MI-RO-VU	JO-JI-TE	FA-GA-PE	GO-RA-RE (frequency -)
Switch phase	PA-TI-VU	TA-MO-TE	DI-PO-BU	TO-KO-LU (frequency +)
	PA-RO-DE	TA-JI-TE	DI-GA-PE	TO-RA-RE (frequency =)
	MI-TI-VU	JO-MO-FU	FA-PO-PE	GO-KO-RE (frequency =)
	MI-RO-DE	JO-JI-FU	FA-GA-BU	GO-RA-LU (frequency -)

Note. The switch phase corresponds to a modification of the last syllable. Notes of frequencies in the XOR Unbalanced condition column corresponds to the frequency distribution of each pseudo-words : + corresponded to a frequency of 0.1, = to 0.0625 (as other conditions) and – to 0.025.

Figure 1

Decrease of response times during the learning phase (10 first blocks) and increase with the switch phase (last block) as a function of condition and predictability (possibility of predicting the present item from the previous items) in Experiment 1.



Note. The error bars correspond to ± 1 standard error.

Figure 2

Structure and transitional probabilities in Experiment 1 for (a) AB Condition, (b) *BC Condition, (c) A*C Condition, and (d) XOR Condition.*

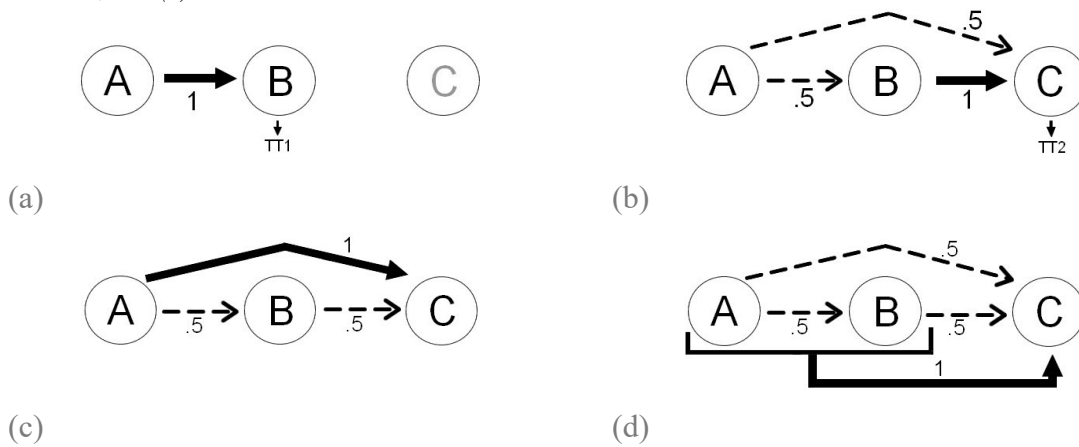
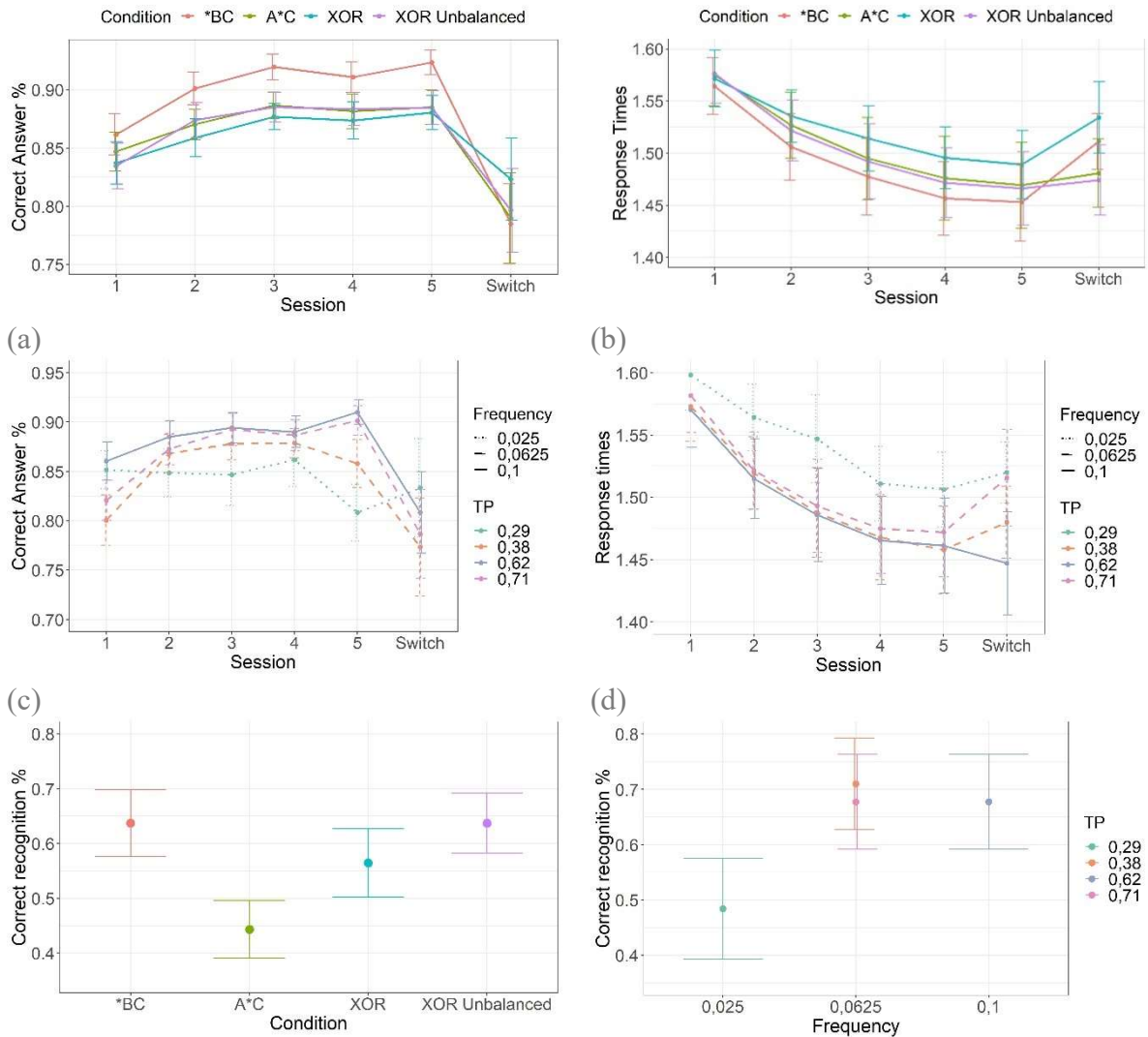


Figure 3

In Experiment 2, increase in performance during the learning phase (5 sessions) and drop in performance with the switch phase (last block) for (a) correct answer rates for each condition, (b) RTs for each condition, (c) correct answer rates for the XOR Unbalanced condition, (d) RTs for the XOR Unbalanced condition, and observed recognition rate in the two-choice recognition task for (e) all conditions and (f) the XOR Unbalanced condition.



(e) Note. The error bars correspond to ± 1 standard error.