Cognition 202 (2020) 104283

Contents lists available at ScienceDirect

Cognition

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

Tuning in to non-adjacencies: Exposure to learnable patterns supports discovering otherwise difficult structures

Martin Zettersten[a,*], Christine E. Potter[b], Jenny R. Saffran[a]

[a] Department of Psychology, University of Wisconsin-Madison, 1202 W Johnson Street, Madison, WI 53706, USA
[b] Department of Psychology, Princeton University, 220 Peretsman Scully Hall, Princeton, NJ 08540, USA

ARTICLE INFO

Keywords:
Non-adjacent dependencies
Language learning
Grammar
Artificial language learning

ABSTRACT

Non-adjacent dependencies are ubiquitous in language, but difficult to learn in artificial language experiments in the lab. Previous research suggests that non-adjacent dependencies are more learnable given structural support in the input – for instance, in the presence of high variability between dependent items. However, not all non-adjacent dependencies occur in supportive contexts. How are such regularities learned? One possibility is that learning one set of non-adjacent dependencies can highlight similar structures in subsequent input, facilitating the acquisition of new non-adjacent dependencies that are otherwise difficult to learn. In three experiments, we show that prior exposure to learnable non-adjacent dependencies - i.e., dependencies presented in a learning context that has been shown to facilitate discovery - improves learning of novel non-adjacent regularities that are typically not detected. These findings demonstrate how the discovery of complex linguistic structures can build on past learning in supportive contexts.

1. Introduction

Non-adjacent dependencies are ubiquitous in language. For instance, English marks number agreement (e.g. *The linguists at the conference are restless*) and aspect (e.g. *People are learning all of the time*) via inflectional morphemes that establish dependencies between distal items. Despite their prevalence in natural languages, non-adjacent dependencies in artificial grammar learning experiments are notoriously difficult to learn, both for adults and infants (e.g., Gómez, 2002; Gonzalez-Gomez & Nazzi, 2012; Newport & Aslin, 2004; Romberg & Saffran, 2013; see Wilson et al., 2018 for a recent review). Given their centrality to language structure, how do we learn non-adjacent dependencies that are not easily detected in speech?

Previous research suggests that the input can be structured to support learners' discovery of non-adjacent regularities. For example, learning can be facilitated simply by increasing exposure (Romberg & Saffran, 2013; Vuong, Meyer, & Christiansen, 2016); additional experience may allow learners more opportunity to uncover patterns. Learning can also be improved when the non-adjacent dependencies are paired with additional cues that highlight their relatedness (e.g., Onnis, Monaghan, Richmond, & Chater, 2005; van den Bos, Christiansen, & Misyak, 2012). For instance, Onnis et al. (2005) found that learners were better able to learn dependencies between phonologically similar syllables, and Newport and Aslin (2004) showed that participants could successfully detect non-adjacent patterns among sets of consonants or vowels, but failed to discover non-adjacent patterns among syllables. Thus, non-adjacent relations seem to be more easily tracked when dependent elements are perceived as similar. Perceptual cues that make relevant items more salient, such as prosody or pauses that mark boundaries in the speech stream, can also boost learning (e.g., Grama, Kerkhoff, & Wijnen, 2016; Peña, Bonatti, Nespor, & Mehler, 2002; Wang & Mintz, 2018), demonstrating that non-adjacent relations can be highlighted in numerous ways.

A particularly powerful factor that can highlight the presence of non-adjacent dependencies is the variability surrounding to-be-learned patterns (Gómez, 2002; Gómez & Maye, 2005). In a classic study by Gómez (2002), participants' learning of non-adjacent regularities improved significantly as the number of unique items that appeared between the dependent elements increased. Variability in the intervening elements affects learning because it can focus attention toward invariant, and hence reliable, structure in the input. With highly variable intermediate elements, learners are better able to detect the reliable associations between non-sequential items, suggesting that surrounding information can help direct learners' attention to non-adjacent regularities.

Learners can also build on past experience with related structures to detect the presence of non-adjacent structures. Previous experience can shape learners' expectations and change the statistical relations that

* Corresponding author.
E-mail address: zettersten@wisc.edu (M. Zettersten).

https://doi.org/10.1016/j.cognition.2020.104283
Received 20 March 2019; Received in revised form 4 March 2020; Accepted 29 March 2020

they track (e.g., LaCross, 2015; Lew-Williams & Saffran, 2012; Potter, Wang, & Saffran, 2017; Wang, Zevin, & Mintz, 2017). For example, experiencing some word categories in adjacent structures subsequently helps learners recognize non-adjacent relations between the same words (Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007). Following experience with associations that are easily learnable, learners may be better able to detect more complex relations (e.g., Elman, 1990; Lai & Poletiek, 2011). Existing native language knowledge can have a particularly powerful impact on the expectations learners form about the structure of upcoming language input. In a recent study, Wang et al. (2017) showed that recent experience with consistent rhythmic patterns embedded in native language structures changes what patterns learners subsequently infer from novel materials. Participants learned non-adjacent dependencies embedded in an artificial language after they were exposed to English phrases that had a matched four-word structure, but not when the two structures were in conflict. This finding is consistent with evidence that infants are better able to discover regularities with a structure that matches their prior experience (Lew-Williams & Saffran, 2012). Together, these studies suggest that learners can use prior experience to improve their learning of non-adjacent dependencies by building on past learning about specific items in simpler contexts or by drawing on knowledge about non-adjacent structures from their first language. However, this leaves open the question of whether learners can discover non-adjacent dependencies *de novo* when the relevant dependencies only appear in non-adjacent relations. When acquiring a novel language, learners must learn new distal grammatical relations that are rarely, if ever, encountered in simpler forms. How might learners break in to learning new non-adjacencies?

In the current work, we investigated whether past distributional learning *itself* may offer a solution to the problem of discovering new non-adjacencies. This explanation focuses on the role of past learning in guiding future learning. If the input is initially structured to support successful non-adjacent dependency learning, this could lead learners to expect to encounter non-adjacent structure in the language. These expectations could subsequently allow them to extract non-adjacent patterns, even in contexts when learning would otherwise be difficult. To test this proposal, we designed a series of experiments in which learners could build on past distributional learning to succeed when faced with a more difficult context for detecting non-adjacent structure. We hypothesized that prior experience with non-adjacent dependencies in the presence of high variability (a context known to support learning; Gómez, 2002; Gómez & Maye, 2005; Plante et al., 2014) would facilitate acquisition of a new set of non-adjacent associations among novel words. In three studies, we tested our hypothesis that experience with one set of non-adjacent dependencies presented in more learnable circumstances would subsequently facilitate learning of a new set of non-adjacent dependencies that learners otherwise struggle to detect. Together, these studies explore how pattern learning in the present builds on pattern learning from the past by testing whether prior experience with readily learnable structures allows difficult linguistic structures to be learned more easily.

2. Experiment 1

Our first study tested whether being pre-exposed to non-adjacent dependencies in a learnable context would aid participants in recognizing novel non-adjacent regularities that are difficult to learn. Learners were tested for their ability to discover the association between the first and third word in three-word sequences (e.g., *pel-kicey-rud*). One group of learners was pre-exposed to a set of artificial sentences that we expected to be learnable based on past work (Gómez, 2002): consistent non-adjacent dependencies with high variability in the intervening elements (Learnable Pre-Exposure Condition). A comparison group was pre-exposed to a set of sentences that was matched in variability but which lacked learnable structure, consisting instead of inconsistent non-adjacent dependencies (Non-Learnable Pre-Exposure Condition). After pre-exposure, all participants were trained on a new set of words organized such that there were consistent non-adjacent dependencies with low variability in the intervening elements – a structure that participants in prior studies consistently struggled to learn (Gómez, 2002). We predicted that participants who had been pre-exposed to learnable non-adjacent relations would more readily detect a new set of non-adjacent dependencies, compared to those who had not. Stimuli, data, and analysis scripts for all experiments are publicly available on the Open Science Framework (https://osf.io/m3wn4/).

2.1. Method

2.1.1. Participants

Sixty-seven students at a large public university in the midwestern United States (37 female; mean age: 18.8 years, $SD = 1.03$; 63 native speakers of English) participated for Introductory Psychology course credit. The sample size was based on previous studies on non-adjacent dependencies, which typically include 12–18 participants per test condition (Frost & Monaghan, 2016; Gómez, 2002; Wang et al., 2017; Wang & Mintz, 2018). We reasoned that increasing our target sample per condition to roughly 30 participants per condition would ensure adequate power based on past research. Participants were randomly assigned to either the Learnable ($n = 32$) or the Non-Learnable ($n = 35$) Pre-Exposure Condition.

2.1.2. Stimuli & design

2.1.2.1. Pre-exposure phase. The stimuli consisted of three-word sequences (e.g., *aXb*) with two monosyllabic words as the first and last elements (e.g., *a* and *b*) and a disyllabic word as the middle element (i.e., an X element). The items in the Pre-Exposure Phase were constructed from 6 monosyllabic nonce words (elements *a-f*: *dak*, *tood*, *feep*, *nov*, *lun*, *kip*) and 24 disyllabic nonce words (X elements: *balip*, *bevit*, *coomo*, *deecha*, *fengle*, *gasser*, *geeble*, *ghope*, *keeno*, *koba*, *lamu*, *loga*, *manu*, *mooper*, *neller*, *riffle*, *rilep*, *roosa*, *skiger*, *suleb*, *tasu*, *toma*, *vulan*, *wasil*). The nonce words were constructed based on items from Gómez (2002) and with the help of a database of pseudowords used in psychological research (Horst & Hout, 2016). Items were selected to follow English phonotactic rules and to include a variety of phonological elements, in particular in word onsets and offsets. Each word was recorded in citation form by a female monolingual speaker of English. Monosyllables and disyllables were each normalized for duration (monosyllables: 655 ms; disyllables: 830 ms) and for average intensity (65–68 dB). The individual items were subsequently concatenated into three-word sentences (*aXb*, *cXd*, etc.; see Fig. 1), with 100 ms of silence between each element within a triplet.

The triplet sequences (e.g., *aXb*) in the Pre-Exposure Phase were either composed of consistent non-adjacent dependencies (Learnable Pre-Exposure Condition) or inconsistent non-adjacent dependencies (Non-Learnable Pre-Exposure Condition; see Fig. 1). In the Learnable Pre-Exposure Condition, there were three non-adjacent dependencies with varying middle (X) elements (*aXb*, *cXd*, *eXf*). We chose a set size of 24 words for the middle X element, since this amount of variability in the intervening word led to the highest learning accuracy in Gómez (2002). We verified that these materials indeed lead participants to successfully learn non-adjacent dependencies in a separate pre-registered experiment, reported in the supplementary materials (see S1 in the supplementary materials for further details). The Non-Learnable Pre-Exposure Condition used the same elements as the Learnable Condition, but recombined them such that there were no consistently predictable (i.e., deterministic) non-adjacent regularities (e.g., elements beginning with the *a* element ended with *b*, *d*, or *f* with equal probability). The triplets were presented one at a time with 750 ms of silence between triplets and in one of two pseudorandomized orders. Orders were created with the constraint that each non-adjacent dependency (i.e., an element with a specific first and last element, e.g.

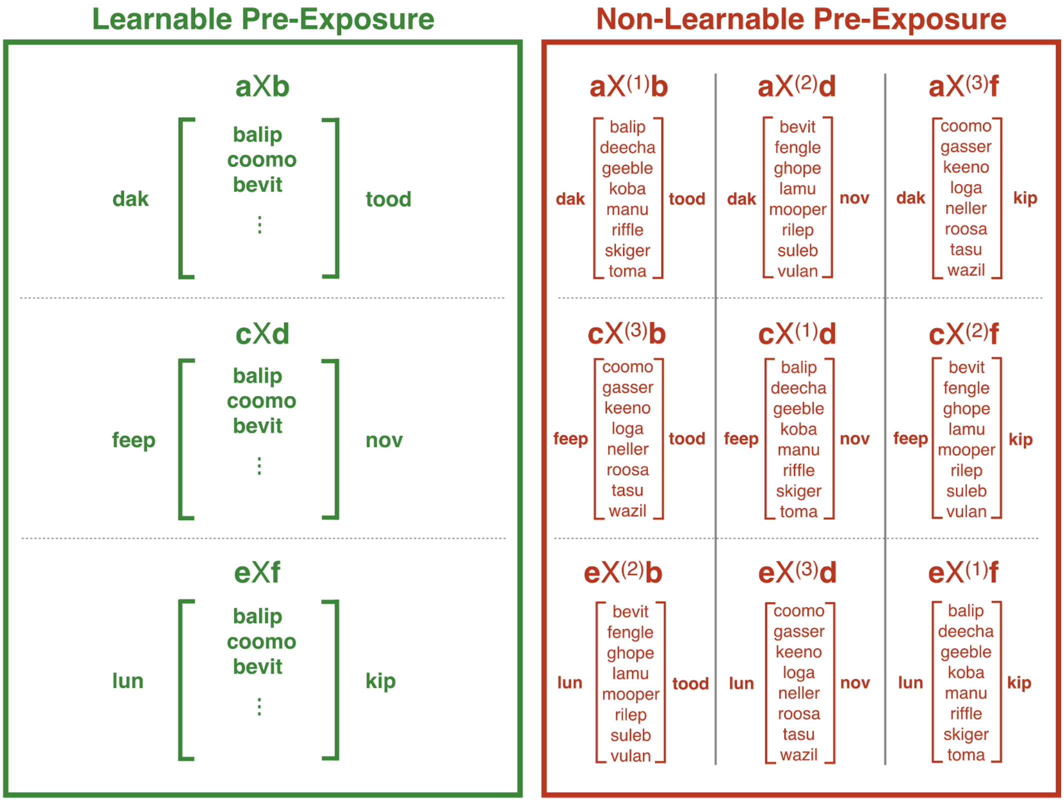

Fig. 1. Pre-Exposure Phase Design in Experiment 1. The full set of 24 X words occurred equally frequently in each condition. In the design of the Non-Learnable Pre-Exposure condition, the 24 medial words were divided into three sets of 8 words each, labeled $X^{(1)}$, $X^{(2)}$, and $X^{(3)}$, to ensure that every X word occurred equally frequently with each initial (*a*, *c*, *e*) and final (*b*, *d*, *f*) word in both Pre-Exposure conditions.

aXb, regardless of the medial element) could occur no more than 3 times in a row. Across the Pre-Exposure phase, participants heard each triplet twice for a total pre-exposure time of 7m25s. This meant that, in the Learnable Pre-Exposure Phase, participants heard each of the three dependencies (*aXb*, *cXd*, *eXf*) $2 * 24 = 48$ times across the Pre-Exposure phase. In the Non-Learnable Pre-Exposure, participants heard each of the nine dependencies ($aX^{(1)}b$, $aX^{(2)}d$, $aX^{(3)}f$, $cX^{(3)}b$, ….) $2 * 8 = 16$ times during the Pre-Exposure phase.

The Learnable Pre-Exposure and the Non-Learnable Pre-Exposure materials share several structural features: for example, the sentences in both conditions follow a monosyllabic word - bisyllabic word - monosyllabic word structure, and items occur consistently in a specific position in the sentence. In principle, these features may also support learning the patterns presented in the Exposure Phase. Crucially, however, the Non-Learnable Pre-Exposure contains no information about non-adjacent dependencies, in the sense of exclusive, deterministic links between specific sentence-initial and sentence-final words.[1] In particular, the structural features contained within the Non-Learnable Pre-Exposure do not distinguish between the test elements that belong to the Exposure language and those that do not (see Test Phase below): for instance, all test items follow the monosyllabic word - bisyllabic word - monosyllabic word structure.

[1] The term "non-learnable" is meant to specifically refer to deterministic non-adjacent links between the first and last words in triplet sequences. The Non-Learnable Pre-Exposure condition contained other dependencies a learner might uncover (e.g., a triplet-specific dependency such as "*feep* followed by *coomo* predicts *tood*"). However, the goal of the design was to ensure that the structure of the Non-Learnable Pre-Exposure was specifically less consistent with the structure of the target non-adjacent dependencies than the Learnable Pre-Exposure.

2.1.2.2. Exposure phase. As in the Pre-Exposure Phase, the Exposure stimuli consisted of three-word sequences (e.g., *gXh*) with monosyllabic words as the first and last elements (e.g., *g* and *h*) and a disyllabic word as the middle element (i.e., an X element). The items in the Exposure Phase were 6 new monosyllabic words (elements *g-l*: *pel*, *rud*, *vot*, *jic*, *bap*, *ghob*) and 3 new disyllabic words (X elements: *kicey*, *puser*, *wadim*). The stimuli were recorded, normalized, and concatenated into three-word sequences in the same manner as the Pre-Exposure stimuli.

In the Exposure Phase (see Fig. 2), participants were randomly assigned to one of two possible languages (L1 or L2) with novel non-adjacent dependencies. In contrast to the Pre-Exposure Phase, the non-adjacent dependencies consisted of only three possible middle elements; prior studies suggest that non-adjacent dependencies with a limited number of middle elements are difficult to learn (e.g., Gómez, 2002). The triplet sequences were presented in one of two pseudo-randomized orders for each language, with the constraint that no triplet could be presented twice in a row, and items with the same non-adjacent dependency (first and third) items could occur no more than three times in a row. Participants heard each of the 9 triplet sequences 24 times across training, for a total Exposure time of 11m7s. Overall, the number of triplet sequences was selected to ensure that the total training time (Pre-Exposure + Exposure = 18m32s) was similar in duration to the 18-min training time from Gómez (2002). In determining the relative durations of the Exposure Phase (60% of the total training time) and the Pre-Exposure phase (40% of the total training time) over the ~18-min training time, we struck a balance between providing participants with sufficient training on the Pre-Exposure items while still allowing participants time to learn the non-adjacent dependencies in the Exposure phase.

2.1.2.3. Test phase. In each of two Test blocks, the *Familiar X Test* block and the *Novel X Test* block, participants were presented with individual

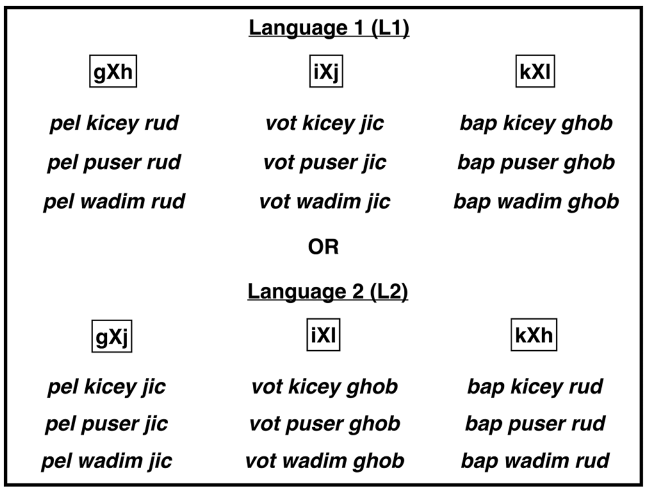

Fig. 2. Exposure phase design (identical across all conditions and experiments).

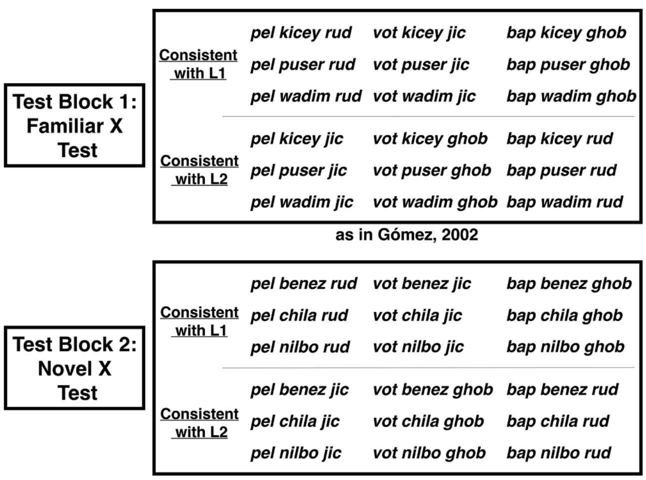

Fig. 3. Test Trial Design (identical across all experiments).

items that either matched or mismatched the Exposure language (see Fig. 3). The Familiar X Test preceded the Novel X Test. Familiar X Test trials matched the test design from Gómez (2002) and were thus always presented first in order to facilitate comparison with past research. During the Familiar X Test block, participants heard 18 triplet sequences: the nine triplets belonging to L1 and the nine triplets belonging to L2. Thus, for each participant, half of the sentences in the Familiar X Test block were sentences presented during the Exposure Phase, while the other half had identical individual words but violated the non-adjacent dependencies from the Exposure Phase. The counterbalancing ensured that items that were familiar in L1 were unfamiliar in L2 and vice versa.

In the Novel X Test block, participants encountered 18 additional sentences (nine consistent with L1 and nine consistent with L2) constructed with three new middle X elements (*benez*, *chila*, *nilbo*) that were not heard during the Exposure Phase (similar items were used by Frost & Monaghan, 2016; Grama et al., 2016). These types of test trials were added as a stricter test that participants had learned the non-adjacent dependencies. On Familiar X trials, participants could conceivably succeed simply by memorizing specific sentences from the Exposure Phase; however, on Novel X trials, participants could only perform accurately if they had encoded the relationship between the first and last words in each triplet. The Novel X sentences were recorded, normalized, and concatenated in the same manner as the three-word sentences in the Pre-Exposure Phase and the Exposure Phase.

2.1.3. Procedure

2.1.3.1. Training phase. The experiment consisted of a two-part Training Phase (Pre-Exposure and Exposure) in which participants listened to the novel language materials through headphones. The Pre-Exposure Phase transitioned seamlessly into the Exposure Phase; the only cue to the transition was the change in the language elements themselves. Participants viewed a series of (unrelated) natural landscape images while listening. Note that the only difference between the Learnable and Non-Learnable Pre-Exposure conditions pertained to the structure of the pre-exposure materials. The Exposure Phase was equivalent across the two conditions.

2.1.3.2. Test phase. The Test Phase consisted of a Familiar X Test block and a Novel X Test block (see Fig. 3). All participants saw the same test items. Participants first judged the 18 triplet sequences from the Familiar X Test block, followed by the 18 triplet sequences from the Novel X Test block. Within each block, the items were presented in random order. For each item, participants were asked to decide whether the sentence matched the word order rules of the language they had just heard, matching procedures from past studies of non-adjacent dependency learning (e.g., Gómez, 2002). Participants were instructed that half of the sentences would match the word order rules of the language and half would not. Participants responded by pressing the "y" key or the "n" key on a keyboard to indicate whether or not sentences matched the word order rules.

2.2. Results

All data and full scripts documenting the data analysis for all experiments are openly available on the Open Science Framework (https://osf.io/m3wn4/). These materials include a walkthrough of all analyses reported in the manuscript, including further modeling details and supplementary analyses described in the Results section (accessible through a web browser at the following link: https://mzettersten.github.io/apg-non-adjacent/data_analysis/APG_analysis.html).

To test the effect of Learnable vs. Non-Learnable pre-exposure on the acquisition of the difficult non-adjacent dependencies in the exposure language, we predicted participants' correct responses across all test trials from Condition (centered; Non-Learnable $= -0.5$, Learnable $= 0.5$) in a logistic mixed-effects model (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). We used the lme4 package version 1.1–21 in R (version 3.6.1) to fit all models (Bates, Mächler, Bolker, & Walker, 2015; R Development Core Team, 2019). We fit the model with the maximal random effects structure, including a by-subject intercept and a by-item random intercept and slope for Condition, and pruned the random effects structure iteratively until arriving at the model with the maximal random effects structure that still allowed the model to converge (Barr, Levy, Scheepers, & Tily, 2013). The final model included a by-subject random intercept and a by-item random slope for condition. Note that in all models reported across these experiments, the parameter estimates and test statistics for the models with the maximal random effects structure and for the final converging models with simplified random effects structure were highly similar and yielded qualitatively equivalent results (see analysis walkthrough documents for details on model outputs).

Collapsing across all test trials, participants in the Learnable Pre-Exposure condition ($M = 62.0\%$, 95% CI $= [55.0\%, 68.9\%]$) were more accurate than participants in the Non-Learnable Pre-Exposure condition ($M = 52.9\%$, 95% CI $= [49.4\%, 56.5\%]$), $b = 0.45$, Wald 95% CI $= [0.08, 0.82]$, $z = 2.39$, $p = .017$ (see Fig. 4). In follow-up analyses investigating whether accuracy differed between test blocks, we found no significant difference in accuracy between Familiar X and Novel X trials ($p = .91$) and no significant interaction between test type and condition ($p = .91$). A similar effect of condition was obtained when Familiar X Test trials ($b = 0.42$, Wald 95% CI $= [0.05, 0.79]$, $z = 2.25$, $p = .024$) and Novel X Test trials ($b = 0.41$, Wald 95%

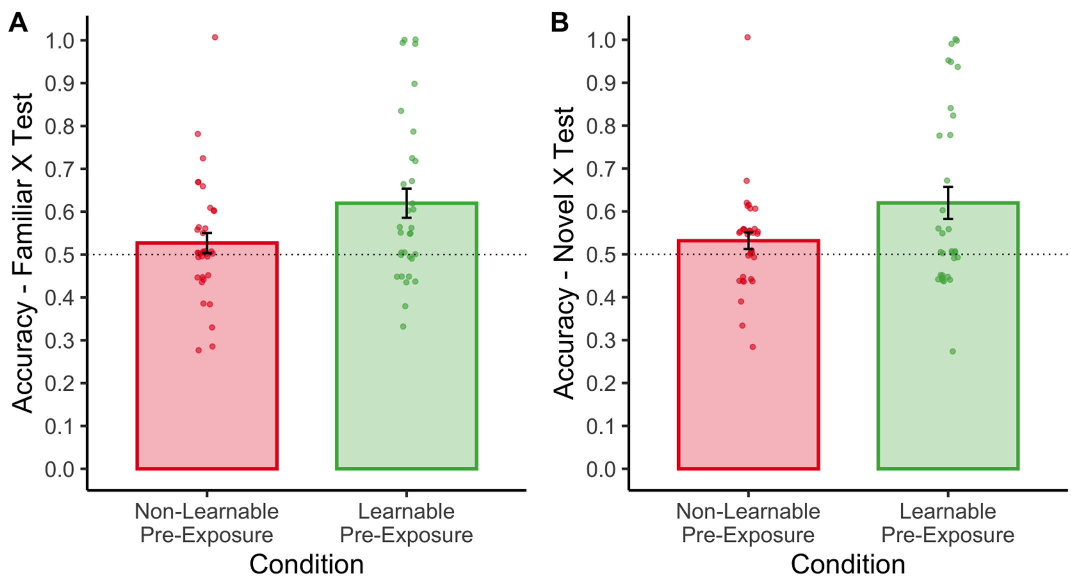

Fig. 4. (A) Familiar X and (B) Novel X Test Accuracy in Experiment 1. Error bars represent +1/ −1 SEs.

CI = [0.04, 0.79], z = 2.16, p = .031) were considered separately. In the supplementary analyses (S2), we report further information on test accuracy within Familiar X and Novel X test trials.

To assess the robustness of the effect to different analytic approaches, we also conducted a signal detection analysis in which we computed the sensitivity index d' and the response bias statistic c for each participant. Following Stanislaw and Todorov (1999), extreme values of 0 and 1 were replaced with $1 / (2 * N)$ and $1 - 1 / (2 * N)$, respectively, where N is the number of total "yes" or "no" trials in a given condition. Participants in the Learnable Pre-Exposure condition (d' = 0.82, 95% CI = [0.33, 1.30]) showed greater sensitivity than participants in the Non-Learnable Pre-Exposure condition (d' = 0.20, 95% CI = [−0.05, 0.45]) in distinguishing items that followed the non-adjacent pattern from items that did not ($t(65)$ = 2.36, p = .02). Participants in both conditions showed a slight bias toward responding "yes", i.e. responding that patterns belonged to the language (Learnable Pre-Exposure condition: c = −0.17, 95% CI = [−0.30, −0.05]; Non-Learnable Pre-Exposure condition: c = −0.30, 95% CI = [−0.48, −0.13]). The magnitude of the response bias did not differ significantly between conditions, $t(65)$ = 1.23, p = .22.

We also investigated the relation between participants' performance on the two test trial types (Familiar X Test vs. Novel X Test). Performance on Familiar X Test trials and Novel X Test trials was correlated in both the Learnable Pre-Exposure condition (r = 0.82, p < .001) and in the Non-Learnable Pre-Exposure condition (r = 0.34, p = .04), though the correlation in the Non-Learnable Pre-Exposure condition was driven by an outlier with perfect performance on both test trial types (Studentized residual t_i = 5.04, Bonferroni-corrected p < .001). When this outlier participant was removed from the analysis, there was no correlation between Familiar X Test and Novel X Test trials in the Non-Learnable Pre-Exposure condition (r = −0.14, p = .44). With or without the inclusion of this outlier participant, there was a significant interaction between test trial type and condition, suggesting that the relation between test trial types was stronger in the Learnable Pre-Exposure condition (outlier included: $t(63)$ = 3.44, p = .001, see Fig. S3 in the supplementary section for a graphical representation of the relationship between accuracy on Familiar X and Novel X trials). Thus, participants who successfully identified non-adjacent patterns that they had heard during training were also more likely to demonstrate generalization of the underlying non-adjacent dependencies, and this relation was stronger in the Learnable Pre-Exposure condition.

2.3. Discussion

Experience with learnable non-adjacent regularities supported participants' ability to learn a new set of non-adjacent regularities that are typically difficult to learn. Participants in the Learnable Pre-Exposure and Non-Learnable Pre-Exposure conditions received *identical* experience with the target language during the Exposure Phase, yet only those participants who had previously heard sequences with learnable non-adjacent dependencies demonstrated evidence of acquiring the novel non-adjacent structures. Accuracy for participants in the Learnable Pre-Exposure condition was comparable to that found in Gómez (2002) for conditions with similar variability in the number of middle elements (Gómez observed 60% accuracy for a set size of two middle elements and 66% accuracy for six middle elements), despite our participants receiving only half of the exposure to the tested non-adjacent relations compared to participants in Gómez (2002). These results are consistent with the hypothesis that prior experience shapes the regularities that learners are able to detect (see e.g., Lew-Williams & Saffran, 2012; Wang et al., 2017). Language learning is cumulative: learners use past experience to constrain their expectations about which statistical regularities to track in novel input (Bates & MacWhinney, 1981; Potter & Lew-Williams, 2019).

By testing participants' ability to generalize to new items, we found strong evidence that participants learned the non-adjacent relations and did not simply memorize the strings that they had encountered before, extending previous research on non-adjacent dependency learning (e.g., Gómez, 2002). Furthermore, the significant correlation between performance on the Familiar X and Novel X Test trials suggests that learning of these associations was relatively robust and could be expressed in multiple ways. The stronger correlation among participants in the Learnable Pre-Exposure condition provides additional evidence that the pre-exposure experience influenced subsequent learning.

As predicted, participants' learning of non-adjacent dependencies was enhanced after the Learnable pre-exposure experience, consistent with our hypothesis that exposure to learnable non-adjacent associations can boost participants' ability to detect new non-adjacent relations. However, there is an alternative explanation for the current results. Because the first and last word in the three-word sentences were not reliably associated in the Non-Learnable pre-exposure, the Pre-Exposure Phase could have drawn participants' attention away from non-adjacent relations and thereby suppressed learning during the Exposure Phase. On this view, rather than enhancing downstream

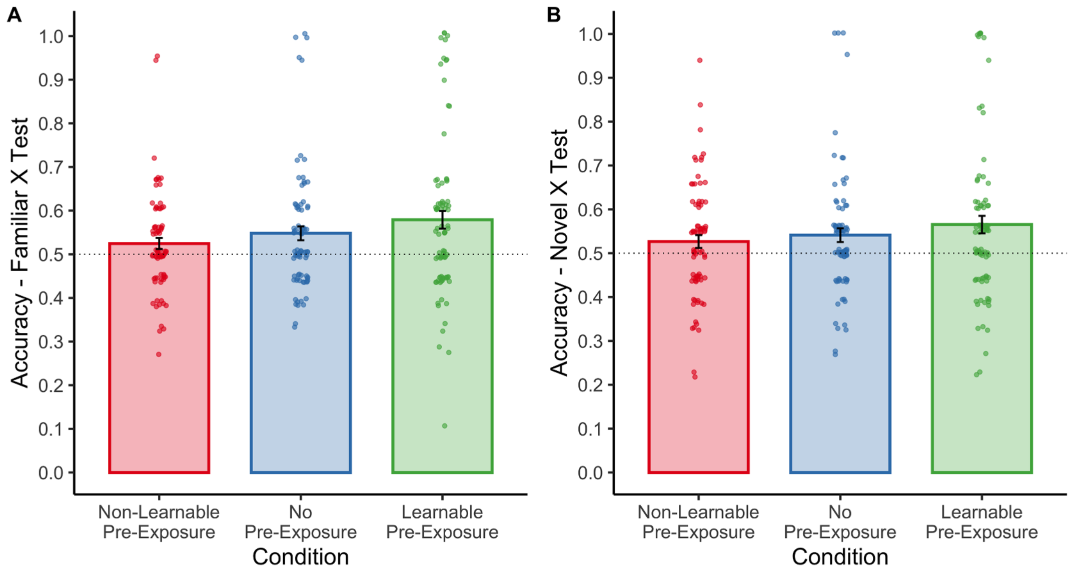

Fig. 5. (A) Familiar X and (B) Novel X Test Accuracy in Experiment 2. Error bars represent +1/ −1 SEs.

learning via exposure to learnable regularities in the Learnable Pre-Exposure condition, downstream learning may have been suppressed by altering the regularities to which learners attended in the Non-Learnable Pre-Exposure condition. A third possibility is that the two conditions both affected learning, but in opposite directions: the Learnable pre-exposure enhanced participants' learning of non-adjacent dependencies, while the Non-Learnable pre-exposure decreased the likelihood that participants would learn novel non-adjacent dependencies. To address these alternatives, we conducted a second experiment with the same conditions as Experiment 1, but with the addition of a baseline condition in which participants were not exposed to either reliable or unreliable non-adjacent dependencies prior to the Exposure Phase. We reasoned that this condition would allow us to estimate the degree to which the pre-exposure manipulations in Experiment 1 aided or suppressed what participants learned about the non-adjacent dependencies in the Exposure Phase.

3. Experiment 2

In Experiment 2, we conducted a replication of Experiment 1 with an additional condition (No Pre-Exposure Condition) in which participants received no pre-exposure experience. We predicted a linear effect across the three conditions, such that performance would be strongest in the Learnable Condition, intermediate in the No Pre-Exposure condition, and weakest in the Non-Learnable Condition, with significant differences between all three conditions. The linear hypothesis and analytic approach were pre-registered on the Open Science Framework (https://osf.io/7ewmc).

3.1. Method

3.1.1. Participants

243 students at a large public university in the midwestern United States (155 female; mean age: 18.5 years, *SD* = 0.86; 203 native speakers of English) participated for Introductory Psychology course credit. Participants were randomly assigned to the Learnable Pre-Exposure (n = 83), the Non-Learnable Pre-Exposure (n = 79), or the No Pre-Exposure Condition (n = 81). A pilot study of the No-Pre-Exposure condition (n = 31) allowed us to estimate the approximate size of the linear effect of condition together with the data from Experiment 1 at $\eta_p = 0.034$. We set a target sample size of 240 participants to ensure that we had over 80% power to detect a linear effect of this size. Three additional participants were tested but excluded due to disruptions of the Training Phase (e.g., falling asleep or leaving the booth to interact with the experimenter).

3.1.2. Stimuli, design & procedure

The stimuli were identical to Experiment 1. The design and procedure for the Non-Learnable and Learnable Pre-Exposure Conditions were identical to Experiment 1. In the No Pre-Exposure Condition, participants did not complete a pre-exposure phase of any kind, instead proceeding straight to the Exposure Phase. As in Experiment 1, the Exposure and Test Phases were identical across all conditions.

3.2. Results

We fit a logistic mixed-effects model to test the linear hypothesis that non-adjacent dependency learning would improve across the three pre-exposure conditions (Non-Learnable < No Pre-Exposure < Learnable), and followed the single contrast approach (Richter, 2015) to analyzing planned contrasts. A statistical approach that tests the residual variance in addition to the planned contrast of interest by including a second orthogonal contrast (Abelson & Prentice, 1997) leads to identical conclusions. We included Condition (coding the planned contrast as Non-Learnable: −0.5, No Pre-Exposure: 0, Learnable: 0.5 to test for a linear increase across conditions) as a fixed effect. The final converging model included a by-subject random intercept. Note that the (non-converging) model with the maximal random effects structure yields equivalent results (see analysis walkthrough for further details). Across all test trials, there was a significant effect of Condition (b = 0.22, Wald 95% CI = [0.03, 0.40], z = 2.31, p = .021), suggesting that there was a linear increase in performance across the three ordered conditions (see Fig. 5). Overall test accuracy increased from the Non-Learnable Pre-Exposure condition (M = 52.6%, 95% CI = [50.5%, 54.7%]) to the No Pre-Exposure condition (M = 54.5%, 95% CI = [51.7%, 57.3%]) to the Learnable Pre-Exposure condition (M = 57.2%, 95% CI = [53.6%, 60.8%]). A similar linear effect of Condition was observed for Familiar X Test trials (b = 0.23, Wald 95% CI = [0.04, 0.43], z = 2.35, p = .019), but this effect was not significant when considering Novel X Test trials alone (b = 0.17, Wald 95% CI = [−0.03, 0.37], z = 1.65, p = .098). However, there was no significant difference in accuracy between Familiar X and Novel X trials (p = .55) and no significant interaction between test type and condition (p = .53).

A similar pattern of results was obtained in a signal detection approach to the analysis, with a linear increase in sensitivity to the non-adjacent relationships from the Non-Learnable Pre-Exposure condition ($d' = 0.16$, 95% CI = [0.03, 0.29]) to the No Pre-Exposure condition ($d' = 0.29$, 95% CI = [0.09, 0.49]) to the Learnable Pre-Exposure condition 3($d' = 0.51$, 95% CI = [0.25, 0.76]), $b = 0.35$, $t(241) = 2.40$, $p = .017$. Participants showed a slight overall bias toward responding that items belonged to the language in all three conditions (Non-Learnable Pre-Exposure condition: $c = -0.30$, 95% CI = [−0.39, −0.22]; No Pre-Exposure condition: $c = -0.28$, 95% CI = [−0.39, −0.18]; Learnable Pre-Exposure condition: $c = -0.20$, 95% CI = [−0.27, −0.12]).

Next, we compared each pair of conditions by conducting pairwise comparisons, using the same modeling approach described above. Participants showed better learning in the Learnable Pre-Exposure condition than in the Non-Learnable Pre-Exposure condition, $b = 0.22$, Wald 95% CI = [0.03, 0.41], $z = 2.26$, $p = .024$, replicating the effect from Experiment 1. However, we found no significant differences between the Learnable Pre-Exposure condition and the No Pre-Exposure condition ($b = 0.14$, Wald 95% CI = [−0.08, 0.36], $z = 1.23$, $p = .22$) or between the No Pre-Exposure condition and the Non-Learnable Pre-Exposure condition ($b = 0.08$, Wald 95% CI = [−0.06, 0.22], $z = 1.09$, $p = .27$). Accuracy reliably differed from chance across all three conditions (Non-Learnable Pre-Exposure: $b = 0.10$, Wald 95% CI = [0.02, 0.19], $z = 2.45$, $p = .014$; No Pre-Exposure: $b = 0.19$, Wald 95% CI = [0.07, 0.32], $z = 3.11$, $p = .002$; Learnable Pre-Exposure: $b = 0.36$, Wald 95% CI = [0.18, 0.55], $z = 3.79$, $p < .001$). Additional pairwise comparisons of pre-exposure conditions considering performance on Familiar X and Novel X trials separately are reported in the supplementary materials (see S2).

As in Experiment 1, we also evaluated the robustness of participants' non-adjacent dependency learning by testing the relation between performance on Familiar X Test trials and Novel X Test trials for participants in the three conditions. If participants are using knowledge of the non-adjacent dependencies to guide their judgments in both the Familiar X and Novel X trials, then accuracy should be highly correlated, since participants who have successfully learned the non-adjacent relations should have similar success on both trial types. Performance on Familiar X Test trials and Novel X Test trials was correlated in the Learnable Pre-Exposure Condition ($r = 0.63$, $p < .001$) and in the No Pre-Exposure Condition ($r = 0.61$, $p < .001$), but not in the Non-Learnable Pre-Exposure Condition ($r = 0.18$, $p = .12$). To test the effect of condition on the relation between performance on the two test trial types, we fit a linear model predicting performance on Novel X Test trials from accuracy on Familiar X Test trials, Condition (contrast coded as Non-Learnable: −0.5, No Pre-Exposure: 0, Learnable: 0.5), and their interaction. There was a significant Familiar X Test accuracy by Condition interaction, suggesting that the strength of the relation between performance on Familiar X Test trials and Novel X Test trials increased across the three (linearly ordered) conditions, $t(239) = 2.45$, $p = .015$ (see Fig. S4 in the Supplementary materials). Thus, in the Learnable Pre-Exposure condition and the No Pre-Exposure condition, participants who performed with higher accuracy on Familiar X test trials also tended to perform better on Novel X trials, while participants in the Non-Learnable Pre-Exposure condition showed a far weaker (if any) relation in their accuracy between the two test blocks.

3.3. Discussion

In Experiment 2, we provided additional evidence that prior experience with reliable or unreliable non-adjacent dependencies can affect subsequent learning. Participants learned novel non-adjacent dependencies better in the Learnable Pre-Exposure condition than in the Non-Learnable Pre-Exposure condition, replicating the results from Experiment 1 (though note that unlike in Experiment 1, there was no significant difference between these two conditions for Novel X trials, see supplementary table S2). Most importantly, the results across our three experimental conditions followed the predicted linear pattern, with accuracy highest in the Learnable Pre-Exposure condition, lowest in the Non-Learnable Pre-Exposure condition, and intermediate in the No Pre-Exposure control condition. The linear increase across the three conditions was significant when considering Familiar X Test trials alone, but only marginal when considering Novel X Test trials alone. This may indicate that the process of generalizing newly learned non-adjacent dependencies to novel sentences is subject to slightly more variability than when identifying previously heard sentences. In general, performance on Familiar X and Novel X test trials was correlated - though the correlation was not significant and weaker in the Non-Learnable Pre-Exposure condition compared to the other two conditions - and the magnitude of the linear effect across conditions was similar for both test trial types, suggesting that performance across the two types of items was highly related. These findings confirm that learning non-adjacent dependencies is sensitive to previous experience with non-adjacent patterns and are consistent with the hypothesis that prior experience has the potential to both facilitate and impair later learning.

Though the main effect was consistent with our linear hypothesis, the individual comparison between the Learnable Pre-Exposure condition did not significantly differ from the No Pre-Exposure condition. Thus, Experiment 2 does not provide conclusive evidence that the higher accuracy observed in the Learnable Pre-Exposure condition (compared to the Non-Learnable Pre-Exposure condition) is truly due to a boost conferred by exposure to learnable non-adjacent dependencies. Importantly, the No Pre-Exposure control condition differed from the two pre-exposure conditions in two ways that may compromise its usefulness as a measure of 'baseline' performance: First, participants in the No Pre-Exposure condition experienced a shorter overall training phase, due to the lack of pre-exposure. Therefore, the slightly improved performance relative to participants in the Non-Learnable Pre-Exposure Condition could be simply due to less fatigue. Second, participants in the No Pre-Exposure condition were exposed to fewer unique language items– they encountered no items outside of the Exposure language. As a consequence, participants may have been less likely to identify test items as "not belonging to the language." Indeed, participants' overall bias for identifying Familiar X Test items as quantified by the response bias statistic c was higher (indicating a higher propensity to respond "yes") in the No Pre-Exposure condition ($c = -0.65$, 95% CI = [−0.75, −0.54]) than in the Learnable Pre-Exposure condition ($c = -0.41$, 95% CI = [−0.50, -0.32]; $t(162) = 3.36$, $p < .001$), though comparable to the response bias in the Non-Learnable Pre-Exposure condition ($c = -0.57$, 95% CI = [−0.68, −0.47]; $t(158) = 0.95$, $p = .34$). These patterns suggest that participants in the No Pre-Exposure condition had different default response behaviors in comparison to the Learnable Pre-Exposure condition, which may be related to the difference in overall language exposure over the course of the experiment.

We therefore designed an additional experiment to compare participants' performance in the Learnable Pre-Exposure condition to a new condition designed to equate learners' total experience with the language materials without biasing them toward or against non-adjacent dependencies. We constructed a condition in which participants received pre-exposure to the same items as in the Learnable Pre-Exposure condition, but with each word presented in isolation rather than in a triplet structure. This design removed any manipulation of participants' expectations regarding connections between the first and last elements in three-item sentences. If participants in the Learnable Pre-Exposure condition have higher test accuracy compared to the new unstructured pre-exposure condition in which words are presented in isolation, this finding would support our interpretation of the results from Experiment 1 and 2 as being at least partially driven by an increase in performance after experience with learnable non-adjacent dependencies.

4. Experiment 3

In Experiment 3, we tested the effect of exposure to learnable non-adjacent dependencies against a new condition (Unstructured Pre-

Exposure Condition) in which total language exposure was equated with materials presented in the Learnable Pre-Exposure Condition. Crucially, the Unstructured Pre-Exposure Condition included a pre-exposure phase consisting of the same words as the pre-exposure in the Learnable Pre-Exposure Condition. However, the words occurred individually in random order, instead of in three-word sentences. Thus, while participants in both conditions heard the same words during the Pre-Exposure Phase, the Unstructured Pre-Exposure Condition did not contain triplet structures. We reasoned that this pre-exposure would provide a baseline condition that should not bias participants' expectations about non-adjacent structure in subsequent items arranged into triplet formats. We predicted that participants in the Learnable Pre-Exposure condition would be more accurate at learning non-adjacent dependencies than participants in the Unstructured Pre-Exposure condition, showing that exposure to learnable non-adjacent dependencies boosts non-adjacent dependency learning. We pre-registered our hypothesis and analytic approach for Experiment 3 on the Open Science Framework (https://osf.io/va657).

4.1. Method

4.1.1. Participants

179 students at a large public university in the midwestern United States (100 female; mean age: 18.8 years, $SD = 1.19$; 157 native speakers of English) participated for Introductory Psychology course credit. We conservatively estimated the effect size for the difference between the Learnable Pre-Exposure condition and our previous control condition with pre-exposure training to be $d = 0.42$. We calculated that a sample size of approximately 180 participants was needed to have 80% power, assuming participants in the Unstructured Pre-Exposure condition would show similar learning to participants in the Non-Learnable Pre-Exposure condition. Participants were randomly assigned to the Learnable Pre-Exposure ($n = 90$) or the Unstructured Pre-Exposure Condition ($n = 89$). One additional participant was tested but excluded due to a technical issue.

4.1.2. Stimuli, design & procedure

The design and procedure for the Learnable Pre-Exposure Condition were identical to Experiments 1 and 2. In the new Unstructured Pre-Exposure Condition, participants listened to pre-exposure materials in which the same individual words occurred with equal frequency as in the Pre-Exposure Phase of the Learnable Pre-Exposure Condition. However, the words were presented in list format in one of two random orders. The duration of silence between each element was 317 ms, such that the total duration of the pre-exposure phase in the new Unstructured Pre-Exposure condition was matched to the pre-exposure phase in the Learnable Pre-Exposure condition (7m25s). We used the word-level recordings from Experiments 1 and 2 to construct the pre-exposure. As in Experiments 1 and 2, the Exposure Phase and subsequent tests for learning of the non-adjacent dependencies were identical in the Learnable and Unstructured Pre-Exposure conditions.

4.2. Results

To test the effect of the Learnable vs. Unstructured pre-exposure materials, we predicted participants' correct responses across all test trials from Condition (centered; Unstructured Pre-Exposure = -0.5, Learnable = 0.5) in a logistic mixed-effects model using the same analytic approach as in Experiment 1. Collapsing across all test trials, participants in the Learnable Pre-Exposure Condition (M = 61.6%, 95% CI = [57.6%, 65.5%]) were more accurate overall than participants in the Unstructured Pre-Exposure Condition (M = 53.6%, 95% CI = [50.9%, 56.3%]), $b = 0.43$, Wald 95% CI = [0.17, 0.68], $z = 3.31$, $p < .001$ (see Fig. 6). There was no significant difference in accuracy between Familiar X and Novel X trials ($p = .60$) and no significant interaction between test type and condition ($p = .87$) (see supplementary section S2 for more detailed information on performance on Familiar X and Novel X test trials). A signal detection approach revealed similar results, with participants showing higher sensitivity to the non-adjacent patterns in the Learnable Pre-Exposure condition ($d' = 0.79$, 95% CI = [0.51, 1.08]) compared to the Unstructured Pre-Exposure condition ($d' = 0.24$, 95% CI = [0.06, 0.42]), $t(177) = 3.23$, $p = .001$. Response bias was comparable between the two conditions, with participants exhibiting a slight tendency toward responding "yes" (Learnable Pre-Exposure: $c = -0.15$, 95% CI = [-0.23, -0.07]; Unstructured Pre-Exposure: $c = -0.16$, 95% CI = [-0.25, -0.08], $t(177) = 0.28$, $p = .78$).

We also investigated the relationship between participants' performance on the two test trial types (Familiar X Test vs. Novel X Test). Performance between Familiar X Test trials and Novel X Test trials was correlated in both the Learnable Pre-Exposure Condition ($r = 0.75$, $p < .001$) and in the Unstructured Pre-Exposure Condition ($r = 0.55$, $p < .001$), though there was a significant interaction between test trial type and condition, suggesting that the relation between Familiar X and Novel X Test performance was stronger in the Learnable Pre-Exposure Condition, $t(175) = 2.88$, $p = .004$ (see Fig. S5 in the supplementary materials). Thus, participants who better recognized the sequences that they had heard during training were also more likely to demonstrate generalization of the underlying non-adjacent dependencies, and this relationship was stronger in the Learnable Pre-Exposure Condition.

4.2.1. Overall analysis across Experiments 1–3

In order to gain a bird's-eye view of the data, we conducted an exploratory analysis testing the effect of condition on test accuracy collapsing across all of the data collected in Experiments 1, 2, and 3 ($N = 489$). Note that this analysis was not pre-registered. We fit a logistic mixed-effects model predicting participants' trial-by-trial test accuracy from condition. Condition was dummy coded, with the Non-Learnable Pre-Exposure condition as the reference level. We first fit a model including the maximal random effects structure (by-participant and by-item random intercepts and a by-item random slopes for condition) and iteratively pruned the random effects structure until convergence was achieved. The final converging model included a by-participant random intercept. All (non-converging) models with more complex random effects structure yield virtually identical parameter estimates and test statistics.

Pre-exposure condition had a strong overall effect on learning, $\chi^2(3) = 24.08$, $p < .001$. We next tested follow-up pairwise comparisons between the Learnable Pre-Exposure condition and the remaining three pre-exposure conditions within the same model by adjusting the reference level in the dummy coded condition variable. Overall, participants in the Learnable Pre-Exposure condition (M = 59.9%, 95% CI = [57.4%, 62.4%]; $d' = 0.68$, 95% CI = [0.50, 0.86]) performed better than participants in the Non-Learnable Pre-Exposure condition (M = 52.7%, 95% CI = [50.9%, 54.5%]; $d' = 0.17$, 95% CI = [0.06, 0.29]; $b = 0.36$, $z = 4.28$, $p < .001$), the Unstructured Pre-Exposure (M = 53.6%, 95% CI = [50.9%, 56.3%]; $d' = 0.24$, 95% CI = [0.06, 0.42]; $b = 0.31$, $z = 3.43$, $p < .001$), and the No Pre-Exposure condition (M = 54.5%, 95% CI = [51.7%, 57.3%]; $d' = 0.29$, 95% CI = [0.09, 0.49]; $b = 0.27$, $z = 2.88$, $p = .004$). Qualitatively similar results are obtained when considering only Familiar X trials or Novel X trials and in analogous signal detection analyses on participants' sensitivity d' (see Fig. S6 and the supplementary analysis walkthrough for further information).

4.3. Discussion

The results of Experiment 3 support the hypothesis that pre-exposure to learnable non-adjacent dependencies can improve the subsequent learning of challenging non-adjacent dependencies, relative to experiencing the same language material in unstructured speech. Participants in the Learnable Pre-Exposure condition were more

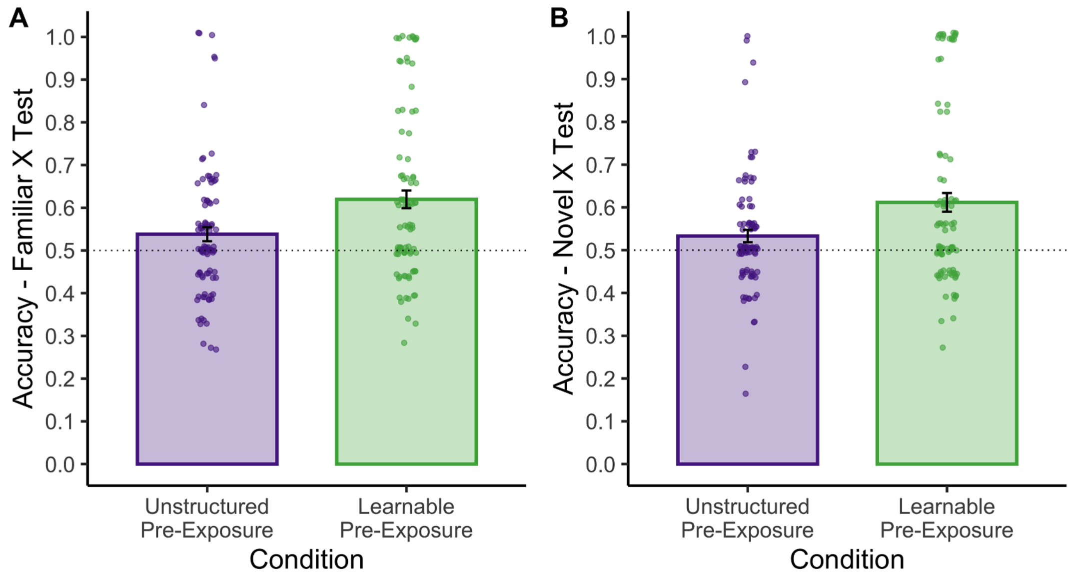

Fig. 6. (A) Familiar X and (B) Novel X Test Accuracy in Experiment 3. Error bars represent +1/ −1 SEs.

successful at learning novel patterns analogous to the non-adjacent patterns from their pre-exposure, compared to participants who heard precisely the same words during pre-exposure without any triplet structure. Combined with the results from Experiments 1 and 2, these findings demonstrate that previous exposure to learnable non-adjacent dependencies increases the likelihood that adult learners will uncover regularities that are otherwise difficult to detect.

5. General discussion

This set of studies investigated a proposal for how distributional learning might build on itself, such that learners develop expectations about linguistic structures that allow them to successfully learn otherwise difficult patterns. When learners were exposed to patterns with learnable non-adjacent dependencies, they were subsequently more successful at learning novel non-adjacent dependencies than if their previous exposure did not include learnable non-adjacent patterns. We tested three separate comparison conditions, where participants received conflicting prior experience (Exp. 1), no prior experience (Exp. 2), and neutral prior experience (Exp. 3), and across all studies found evidence for enhanced learning following exposure to learnable regularities. We also found no evidence to suggest that, given such experience, participants performed differently when tested on sentences with familiar medial elements (Familiar X Test) and when they were required to generalize to sentences with novel medial elements (Novel X Test). Thus, we have provided evidence that experience can support the subsequent learning of non-adjacent patterns, and we suggest that one way that learners may more readily discover challenging novel regularities in language is to make use of knowledge previously abstracted from similar regularities.

The increase in accuracy for participants in the Learnable Pre-Exposure compared to the other three conditions was modest: across all experiments, learners in the Learnable Pre-Exposure condition had 5–7% higher accuracy on average than participants in the other three experimental conditions without favorable pre-exposure input. This difference in group accuracy corresponds to a small to moderate effect size (comparison of the Learnable Pre-Exposure condition to the No Pre-Exposure condition: Cohen's $d = 0.32$; Unstructured Pre-Exposure: $d = 0.38$; Non-Learnable Pre-Exposure: $d = 0.46$). While a difference of 5–7% is relatively small in terms of absolute accuracy increase, a boost in accuracy of this magnitude is consequential in the context of the difficulty of learning non-adjacent structures. Given that participants performed only slightly above chance in the exposure phase without favorable pre-exposure (M = 52%–55% in the three non-favorable conditions), a small boost in accuracy indicates that learners are more reliably uncovering structure they might otherwise not learn at all. Even a short amount of prior experience (~8 min) with non-adjacent structure in more learnable circumstances can provide an advantage in uncovering similar structures in more challenging contexts.

Adult participants' improved performance after exposure to related structures is consistent with evidence from both infants and adults showing that prior experience influences learners' expectations about the structure of novel linguistic materials (e.g., Lew-Williams & Saffran, 2012; Thiessen & Saffran, 2007; Wang et al., 2017). It has been suggested that past experience may guide learners to identify similar structures to those stored in memory (Thiessen, Kronstein, & Hufnagle, 2013). On this view, once participants in the Learnable Pre-Exposure condition had learned the associations between the monosyllabic first and third elements during the pre-exposure phase, this learning experience allowed them to better detect associations between new monosyllabic items. This explanation suggests that learners should be most likely to generalize across tokens that are highly similar, as observed in a number of prior studies (Christiansen & Conway, 2006; Gebhart, Newport, & Aslin, 2009; Newport & Aslin, 2004; Seidl-Rathkopf, Turk-Browne, & Kastner, 2015), but they might not be able to draw on experience with related structures when the target items are perceptually unrelated.

Another possibility is that experience with learnable dependencies between non-adjacent items draws attention to subsequent non-adjacent relations. Shifting attention toward non-adjacent dependencies can support learners' ability to detect novel regularities (Pacton & Perruchet, 2008). In addition, past studies have demonstrated that learnable patterns may automatically attract attention, and learners are more inclined to attend to elements and positions that have previously included predictable items (e.g., Gerken, Balcomb, & Minton, 2011; Turk-Browne, Scholl, Johnson, & Chun, 2010; Zhao, Al-Aidroos, & Turk-Browne, 2013). Therefore, prior linguistic experience may encourage learners to detect regularities that occur in similar contexts as patterns that they have learned in the past, even when the items are novel.

In the current design, we chose to highlight non-adjacent regularities by introducing intervening variability, but a number of different cues have also been shown to support learning (Grama et al., 2016; Peña et al., 2002; see Wilson et al., 2018 for a review). Other types of pre-exposure materials may be equally helpful in preparing learners to discover the relations in the exposure phase. For example, we could have increased the salience of the non-adjacent items in the pre-exposure phase, such as introducing phonological cues to underscore the associations between the critical elements (Gervain & Endress, 2017; van den Bos et al., 2012). It may be that any cue that facilitates learning of the initial relations has the potential to bolster later learning, but additional studies will be needed to test whether variability is a particularly powerful cue in learning, or if other cues are similarly advantageous.

Overall, our results are consistent with the view that adults' language learning is constrained by prior language experience and knowledge (Bates & MacWhinney, 1981; Seidenberg & Zevin, 2006). Learners do not simply blindly track statistical relations in their input, but instead use their experience to determine which regularities are likely to be most meaningful (e.g., Frank & Tenenbaum, 2011; Lew-Williams & Saffran, 2012; Mintz, 2002; Potter & Lew-Williams, 2019). In these studies, participants who had experience with reliable associations were then able to detect patterns in novel materials. This pattern of performance is consistent with the view that learners gradually accumulate knowledge over time that prepares them to learn complexities found in natural language (e.g., Seidenberg, MacDonald, & Saffran, 2002; Thiessen & Saffran, 2007).

These cumulative effects of learning are especially apparent in adults (who already have vast experience with their native language) that are learning a second language. Adult second language learners are better able to acquire constructions that are consistent with the regularities of their native language (LaCross, 2015), providing additional support for the view that some of adults' difficulties in learning a second language may be attributable to biases derived from their experience with their first language (e.g., Marchman, 1993; Seidenberg & Zevin, 2006). For example, native English speakers have significant difficulty with grammatical gender and often struggle to assign the correct article to a noun, but learners whose first language uses gender are more successful (Grüter, Lew-Williams, & Fernald, 2012; Guillelmon & Grosjean, 2001; Sabourin, Stowe, & de Haan, 2006). Lifelong language experience encourages learners to pay attention to or ignore some structures (such as associations between articles and nouns) rather than others (see also, e.g., Siegelman, Bogaerts, Elazar, Arciuli, & Frost, 2018).

The question of how past experience affects subsequent language learning opens a number of future directions. One particularly intriguing direction is studying individual differences in how people extract regularities from language patterns. Participants showed substantial variability in their ability to learn the non-adjacent dependencies across all experiments and conditions (see the spread in average accuracy in Figs. 4, 5, and 6), with a large set of participants showing chance or near-chance performance and a subset of participants showing virtually perfect performance. While some research has investigated individual differences in statistical learning and their relation to language processing and knowledge (e.g., Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010), we still know little about what underlies differences in learning statistical patterns such as non-adjacent dependencies between learners. Given the role of past language exposure in shaping later learning, a particularly fruitful direction for future research may be to explore the degree to which particular linguistic experience predicts individual differences in non-adjacent dependency learning.

Another question left open in the current work is the relationship between the learnability of the target structure and the degree to which learning can be improved with informative initial experience. While learning the non-adjacent dependencies in low variability contexts is difficult, participants still performed slightly above chance in discovering non-adjacent relations absent any pre-exposure (M = 54.5% in the No Pre-Exposure condition of Experiment 2). In other words, the non-adjacent associations presented during the Exposure Phase were to some extent learnable. It is possible that some degree of learnability is crucial for learning to be malleable to informative past experience, as observed in the current studies. However, it is inherently difficult to show that some language structure is truly unlearnable (e.g., Chater & Manning, 2006; Regier & Gahl, 2004; St Clair, Monaghan, & Ramscar, 2009), since absence of evidence (for learnability) is not equivalent to evidence of absence (i.e., that the structure is unlearnable). In fact, learning non-adjacent relations with similarly small set sizes in the middle word has sometimes been described as not learnable or only marginally learnable absent additional cues, based on the absence of above-chance performance (Gómez, 2002; Newport & Aslin, 2004; von Koss Torkildsen, Dailey, Aguilar, Gómez, & Plante, 2013; though studies also sometimes find learning even with low variability in the middle elements, e.g., Vuong et al., 2016), which is not surprising given the small learning effect observed in our sample. Future work could address this question by systematically manipulating the difficulty of the to-be-learned language structure to test whether more difficult structures are more or less amenable to supportive past experience.

To conclude, we provide evidence that relatively brief experience can have substantive consequences for the types of patterns to which learners are sensitive. Non-adjacent dependencies may be relatively difficult to learn in isolation, but previous language experience and accumulated knowledge can make these regularities easier (or in some cases, harder) to learn by helping learners recognize and attend to important patterns in their language input. Past language experience matters because it sets the stage for later learning, building a scaffold to acquiring otherwise difficult patterns.

Author contributions

All authors developed the study concept and design. Data collection and data analysis were performed by MZ. All authors contributed to the interpretation of the data and wrote the manuscript.

Acknowledgements

This research was supported by NSF-GRFP DGE-1747503 awarded to MZ, and grants from the NICHD to JRS (R37HD037466), CEP (F32 HD093139), and the Waisman Center (U54 HD090256). We thank Jill Lany for helpful comments on an earlier draft, and Emily Cummings, Grace McCune, Lauren Silber, and Amy So for aiding in data collection.

Appendix A. Supplementary Materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104283.

References

Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypotheses. *Psychological Methods, 2*(4), 315–328. https://doi.org/10.1037/1082-989X.2.4.315.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bates, E. A., & MacWhinney, B. (1981). Second language acquisition from a functionalist perspective: Pragmatic, semantic, and perceptual strategies. In H. Winitz (Ed.). *Native language and foreign language acquisition. Annals of the New York Academy of Sciences, no. 379* (pp. 190–214). New York: New York Academy of Sciences.

van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language, 67*(4), 507–520. https://doi.org/10.1016/j.jml.2012.07.008.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, 10*(7), 335–344. https://doi.org/10.1016/j.tics.2006.05.006.

Christiansen, M. H., & Conway, C. M. (2006). Statistical learning within and between

modalities: Pitting abstract against stimulus-specific representations. *Psychological Science, 17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x.
Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1.
Frank, M. C., & Tenenbaum, J. B. (2011). Three ideal observer models for rule learning in simple languages. *Cognition, 120*(3), 360–371. https://doi.org/10.1016/j.cognition.2010.10.005.
Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition, 147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010.
Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin and Review, 16*(3), 486–490. https://doi.org/10.3758/PBR.16.3.486.Statistical.
Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid "labouring in vain" by attending more to learnable than unlearnable linguistic patterns. *Developmental Science, 14*(5), 972–979. https://doi.org/10.1111/j.1467-7687.2011.01046.x.
Gervain, J., & Endress, A. D. (2017). Learning multiple rules simultaneously: Affixes are more salient than reduplications. *Memory and Cognition, 45*(3), 508–527. https://doi.org/10.3758/s13421-016-0669-9.
Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476.
Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183–206. https://doi.org/10.1207/s15327078in0702_4.
Gonzalez-Gomez, N., & Nazzi, T. (2012). Acquisition of nonadjacent phonological dependencies in the native language during the first year of life. *Infancy, 17*(5), 498–524. https://doi.org/10.1111/j.1532-7078.2011.00104.x.
Grama, I. C., Kerkhoff, A., & Wijnen, F. (2016). Gleaning structure from sound: The role of prosodic contrast in learning non-adjacent dependencies. *Journal of Psycholinguistic Research, 45*(6), 1427–1449. https://doi.org/10.1007/s10936-016-9412-8.
Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research, 28*(2), 191–215. https://doi.org/10.1177/0267658312437990.
Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition, 29*(3), 503–511. https://doi.org/10.3758/BF03196401.
Horst, J. S., & Hout, M. C. (2016). The novel object and unusual name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007.
von Koss Torkildsen, J., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research, 56*(2), 618–629. https://doi.org/10.1044/1092-4388(2012/11-0125).
LaCross, A. (2015). Khalkha Mongolian speakers' vowel bias: L1 influences on the acquisition of non-adjacent vocalic dependencies. *Language, Cognition and Neuroscience, 30*(9), 1033–1047. https://doi.org/10.1080/23273798.2014.915976.
Lai, J., & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition, 118*(2), 265–273. https://doi.org/10.1016/j.cognition.2010.11.011.
Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science, 19*(12), 1247–1252. https://doi.org/10.1111/j.1467-9280.2008.02233.x.
Lany, J., Gómez, R. L., & Gerken, L. (2007). The role of prior experience in language acquisition. *Cognitive Science, 31*, 481–507. https://doi.org/10.1080/15326900701326584.
Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition, 122*(2), 241–246. https://doi.org/10.1016/j.cognition.2011.10.007.
Marchman, V. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5*(2), 215–234. https://doi.org/10.1162/jocn.1993.5.2.215.
Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition, 30*(5), 678–686. https://doi.org/10.3758/BF03196424.
Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning, 62*(1), 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x.
Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1*(31), 1–9. https://doi.org/10.3389/fpsyg.2010.00031.
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2.
Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53*(2), 225–237. https://doi.org/10.1016/j.jml.2005.02.011.
Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34*(1), 80–96. https://doi.org/10.1037/0278-7393.34.1.80.
Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604–607. https://doi.org/10.1126/science.1072901.
Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., ... Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology, 23*(4), 530–545. https://doi.org/10.1044/2014_AJSLP-13-0038.
Potter, C. E., & Lew-Williams, C. (2019). Infants' selective use of reliable cues in multidimensional language input. *Developmental Psychology, 55*(1), 1–8. https://doi.org/10.1037/dev0000610.
Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive Science, 41*, 913–927. https://doi.org/10.1111/cogs.12473.
R Development Core Team (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.
Regier, T., & Gahl, S. (2004). Learning the unlearnable: The role of missing evidence. *Cognition, 93*(2), 147–155. https://doi.org/10.1016/j.cognition.2003.12.003.
Richter, M. (2015). Residual tests in the analysis of planned contrasts: Problems and solutions. *Psychological Methods, 21*(1), 112–120. https://doi.org/10.1037/met0000044.
Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science, 37*(7), 1290–1320. https://doi.org/10.1111/cogs.12050.
Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22*(1), 1–29. https://doi.org/10.1191/0267658306sr259oa.
Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science, 298*(5593), 553–554. https://doi.org/10.1126/science.1078094.
Seidenberg, M. S., & Zevin, J. D. (2006). Connectionist models in developmental cognitive neuroscience: Critical periods and the paradox of success. In Y. Munakata, & M. H. Johnson (Eds.). *Attention and performance XXI: Processes of change in brain and cognitive development* (pp. 585–612). Oxford: Oxford University Press.
Seidl-Rathkopf, K. N., Turk-Browne, N. B., & Kastner, S. (2015). Automatic guidance of attention during real-world visual search. *Attention, Perception, and Psychophysics, 77*(6), 1881–1895. https://doi.org/10.3758/s13414-015-0903-8.
Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition, 177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011.
St Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science, 33*(7), 1317–1329. https://doi.org/10.1111/j.1551-6709.2009.01065.x.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704.
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), 792–814. https://doi.org/10.1037/a0030801.
Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development, 3*(1), 73–100. https://doi.org/10.1080/15475440709337001.
Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience, 30*(33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010.
Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent statistical learning of adjacent and nonadjacent dependencies. *Language Learning, 66*(1), 8–30. https://doi.org/10.1111/lang.12137.
Wang, F. H., & Mintz, T. H. (2018). Learning nonadjacent dependencies embedded in sentences of an artificial language: When learning breaks down. *Journal of Experimental Psychology: Learning Memory and Cognition, 44*(4), 604–614. https://doi.org/10.1037/xlm0000483.
Wang, F. H., Zevin, J. D., & Mintz, T. H. (2017). Top-down structure influences learning of nonadjacent dependencies in an artificial language. *Journal of Experimental Psychology: General, 146*(12), 1738–1748. https://doi.org/10.1037/xge0000384.
Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., ... Rey, A. (2018). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science,* 1–16. https://doi.org/10.1111/tops.12381.
Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science, 24*(5), 667–677. https://doi.org/10.1177/0956797612460407.