EMPIRICAL STUDY

Concurrent Statistical Learning of Adjacent and Nonadjacent Dependencies

Loan C. Vuong,[a] Antje S. Meyer,[a,b] and Morten H. Christiansen[c]

[a]Max Planck Institute for Psycholinguistics, [b]Radboud University, and [c]Cornell University

When children learn their native language, they have to deal with a confusing array of dependencies between various elements in an utterance. The dependent elements may be adjacent to one another or separated by intervening material. Prior studies suggest that nonadjacent dependencies are hard to learn when the intervening material has little variability, which may be due to a trade-off between adjacent and nonadjacent learning. In this study, we investigate the statistical learning of adjacent and nonadjacent dependencies under low intervening variability using a modified serial reaction time (SRT) task. Young adults were trained on mixed sets of materials comprising equally probable adjacent and nonadjacent dependencies. Offline tests administered after training showed better performance for adjacent than nonadjacent dependencies. However, online SRT data indicated that the participants developed sensitivity to both types of dependencies during training, with no significant differences between dependency types. The results demonstrate the value of online measures of learning and suggest that adjacent and nonadjacent learning can occur together even when there is low variability in the intervening material.

Keywords statistical learning; artificial grammar learning; serial reaction time; nonadjacent dependencies

Introduction

It is generally assumed that statistical learning, a mechanism that encodes statistical regularities in the environment, plays a role in language acquisition (see Aslin & Newport, 2009; Seidenberg, 1997). An extensive body of work has shown that adults as well as infants can readily pick up on regularities among adjacent elements, such as among adjacent speech sounds or visual shapes (e.g., Bulf, Johnson, & Valenza, 2011; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996). This sensitivity to adjacent dependencies has been found to be present even in newborns (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009).

Beyond adjacencies, natural languages also involve a variety of nonadjacent dependencies, in morphosyntax—as in nonadjacent verb patterns such as *is*/*are* verb-*ing*, and sentence-level syntax—as in embedded sentence structures such as *The man the boy saw kissed the woman*. Therefore, an important research topic is whether statistical learning is possible when the dependent elements are nonadjacent. As adjacent and nonadjacent dependencies coexist in natural languages, a further issue is whether statistical learning of both adjacent and nonadjacent dependencies can occur together. Answering these questions is essential to determine the importance of statistical learning in natural language acquisition.

Concurrent Statistical Learning of Adjacent and Nonadjacent Dependencies

The empirical findings to date suggest that statistical learning of nonadjacent regularities is possible under favorable conditions (e.g., Creel, Newport, & Aslin, 2004; Gebhart, Newport, & Aslin, 2009; Gómez, 2002; Gómez & Maye, 2005; Newport & Aslin, 2004; Onnis, Christiansen, Chater, & Gómez, 2003; van den Bos, Christiansen, & Misyak, 2012). Nonadjacent learning can, for instance, readily occur when perceptual cues that group together the nonadjacent elements are available (Creel et al., 2004; Gebhart et al., 2009; Newport & Aslin, 2004; van den Bos et al., 2012), or when learners are preexposed to nonadjacent pairs in an adjacent structure prior to being exposed to the same pairs in a nonadjacent structure (Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007). Notably, it has been found that the variability of the materials intervening between the nonadjacent elements places an important constraint on nonadjacent learning (Gómez, 2002; Gómez & Maye, 2005; Onnis et al., 2003). Gómez (2002) trained participants on an artificial language that consisted of three-element spoken sequences with deterministic nonadjacent dependencies

between the first and third elements (e.g., A x B and C x D, nonadjacent probability $p = 1.0$). The variability of the intervening element was manipulated by varying the set size of the second elements in four between-participant conditions (x = 2, 6, 12, or 24 exemplars). The manipulation resulted in lower intervening variability in the conditions with smaller set sizes and higher intervening variability in those with larger set sizes. Participants listened to the sequences for approximately 20 minutes, after which they made grammaticality judgments on strings that either followed trained patterns (e.g., AxB and CxD) or deviated from trained patterns in the third element position (e.g., AxD and CxB). Gómez found that learners accepted the grammatical strings at high rates across all four conditions (mean acceptance rates 77% to 100%), but those trained under larger intervening set sizes were less likely to erroneously accept the ungrammatical strings, relative to those trained under smaller intervening set sizes (mean acceptance = 20%, 47%, 54%, and 68% for x = 24, 12, 6, and 2, respectively). It thus appears that nonadjacent dependencies can be learned readily when the adjacent elements are highly variable, but perhaps not when these elements have less variability. Interestingly, individuals with specific language impairment have been found unable to use variability in statistical learning of nonadjacent dependencies (Hsu, Tomblin, & Christiansen, 2014).

Gómez's (2002) findings raise the possibility that nonadjacent learning may be hindered by adjacent learning. In particular, the poorer judgments with increasing strength of the adjacent statistics may suggest a "tendency to focus on adjacent dependency," which may be "fundamental in learning and may even be the default" (Gómez, 2002, pp. 434–435). According to this possibility, learners will focus on adjacent relations upon encountering exploitable adjacencies, eventually succeeding in adjacent learning but having difficulty with or even failing in nonadjacent learning. However, upon encountering uninformative adjacent dependencies, learners may switch attention to informative nonadjacent relations and consequently succeed in acquiring nonadjacent dependencies. If true, this would suggest a severe limitation in learners' ability to track adjacent and nonadjacent statistics concurrently (see also Pacton & Perruchet, 2008).

Romberg and Saffran (2013) investigated whether adjacent and nonadjacent dependencies can be learned concurrently or whether there may be trade-offs between adjacent and nonadjacent learning (see also Vuong, Meyer, & Christiansen, 2011). Similar to Gómez (2002), they trained adult learners on an AxB grammar with deterministic nonadjacent dependencies between the first and third elements (three A-B frames, $p = 1.0$ each). Embedded within the nonadjacent frames were second element exemplars (x = 12) that resulted in high intervening variability for nonadjacent learning. Adjacent relations (Ax or

xB) were probabilistic and less informative than the nonadjacent relations (highest adjacent $p = .80$ and $p = .67$ in Experiments 1 and 2, respectively). Different groups of learners listened to the spoken sequences at exposure durations that ranged from 5 to 20 minutes. At test, all learners performed two judgment tasks in one of two testing orders (adjacent test first or nonadjacent test first). The adjacent test involved discriminating between strings with legal nonadjacent frames surrounding a trained or untrained second element for the frames. The nonadjacent test involved discriminating between strings with trained or untrained nonadjacent frames while holding constant the likelihood of the adjacent relations within the frames. The main results showed that learners performed above chance on both tests with no significant differences between tests across exposure durations. This suggests that adjacent and nonadjacent dependencies can be learned rapidly together, at least given high intervening variability.

The Present Study

The present study sought to provide a stronger test of concurrent adjacent and nonadjacent learning, by investigating learning under low intervening variability. As reviewed above, it is under this condition that nonadjacent learning has been found to be more difficult than under high intervening variability (e.g., Gómez, 2002). We therefore used a smaller intervening set size of four exemplars, as compared to a set size of 12 intervening exemplars in Romberg and Saffran (2013). Furthermore, Romberg and Saffran's training materials featured deterministic nonadjacent dependencies but probabilistic adjacent dependencies within the same set of materials. Their materials were therefore biased toward the acquisition of nonadjacent dependencies (see also De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Lévi, 2007). By contrast, there was no such bias in our study, as adjacent and nonadjacent statistics were matched by implementing probabilistic adjacent and probabilistic nonadjacent dependencies in separate sets of materials that were randomly intermixed during training ($p = .50$ each). Using different materials across the dependency conditions allowed us to observe nonadjacent learning when adjacent pairs were made uninformative not only during the testing phase but also during the learning phase, ruling out the possibility that nonadjacent performance may be mediated by adjacent pairs in either of the phases.

Similar to Romberg and Saffran (2013), we presented adult learners with three-element sequences during training and asked them to perform offline tests at the end of training (see also Gómez, 2002). However, we greatly extended the training given to participants. Compared to natural learning situations, the duration of exposure given in artificial grammar learning studies tends

to be much reduced—typically around 20 minutes in studies on nonadjacent learning (e.g., Creel et al., 2004; Gómez, 2002; Gómez & Maye, 2005; Onnis et al., 2003; Romberg & Saffran, 2013; van den Bos et al., 2012). Failure to observe learning in the laboratory might stem from the limited learning opportunity given to learners rather than from inherent limits on learning. A recent study by van den Bos and colleagues (2012) has examined learning of probabilistic nonadjacent dependencies within this time frame. No learning was found unless visual or phonological grouping cues were present. In this study, we lengthened each training session to approximately 1 hour, and (more critically perhaps) there were three sessions on successive days, allowing for potential memory consolidation of dependency knowledge to take place (e.g., see Durrant, Taylor, Cairney, & Lewis, 2011).

Finally, another novel aspect of the current study concerned the assessment of learning. In addition to offline testing after training, we tracked each participant's performance online using a modified serial reaction time (SRT) task (Misyak, Christiansen, & Tomblin, 2010a). The SRT task (Nissen & Bullemer, 1987) is often used in studies of sequence learning, which measures reaction times (RTs) to sequence elements in structured and random trials online during learning. Pattern-specific learning is revealed as RTs to structured blocks decrease relative to RTs in random blocks. The SRT task, which does not require participants to reflect upon trained patterns, can be distinguished from offline tests that require participants to make explicit decisions about trained patterns, such as predicting the next element in a sequence or making judgments about the grammaticality of test sequences. Prior studies have shown that results obtained via the SRT task may be dissociated from those obtained via offline tasks (e.g., Cherry & Stadler, 1995; Gaillard, Destrebecqz, Michiels, & Cleeremans, 2009; Howard & Howard, 1989, 1992; Jiménez, Méndez, & Cleeremans, 1996; Willingham, Greeley, & Bardone, 1993; see Dienes & Berry, 1997, for a review).

As noted above, Romberg and Saffran (2013) measured adjacent and nonadjacent learning using offline judgments. In their study, taking the adjacent test first was found to result in worse performance on nonadjacent than adjacent dependencies in the first experiment, and taking the nonadjacent test first resulted in worse adjacent than nonadjacent performance in the second experiment (though the reverse patterns did not obtain within the same experiments). Because only offline judgments were available, it is not possible to tell whether the patterns observed reflect differences in learning or whether they might arise during offline testing. In our study, therefore, we included an online SRT measure in addition to two offline ones, a prediction and a grammaticality

judgment task (see the Method section). As grammaticality judgments may involve meta-cognition to a greater extent than prediction, it was administered last during offline testing. Finally, and potentially relevant for applications to language learning, the modified SRT task used here has been shown to provide a sensitive indicator of learning and correlate with online language processing ability in earlier studies (Misyak & Christiansen, 2010; Misyak et al., 2010a; Misyak, Christiansen, & Tomblin, 2010b).

To summarize, our study investigated concurrent learning of probabilistic adjacent and probabilistic nonadjacent dependencies under low intervening variability. The adjacent and nonadjacent dependencies were featured in separate sets of materials and were matched in probabilities. Our participants were exposed to the materials in an extended SRT training phase, which provides an online measure of learning. Dependency learning was additionally tested using two offline tasks administered after the last training session. If statistical learning of adjacent and nonadjacent dependencies can occur together, despite low intervening variability, learners should show sensitivity to both types of dependencies, at least in the online SRT measure. The concurrent learning may or may not similarly manifest in the offline measures, as offline performance is subject to additional influences from meta-cognitive decision making. Alternatively, if statistical learning is biased toward adjacent dependencies to the exclusion of nonadjacent dependencies, then upon encountering the exploitable adjacencies learners should show sensitivity to the adjacent dependencies but not to the nonadjacent dependencies. Such a bias, if true, should be evident in the online measure, and perhaps also in the offline ones, as suggested by previous findings (e.g., Gómez, 2002).

Method
Participants
Forty-five Dutch native speakers (14 males, M_{age} = 21.3 years, SD = 2.9) participated in the experiment. They were paid €24 for the three 1-hour sessions held on successive days.

Training Materials
Training materials consisted of spoken three-element sequences of Dutch pseudowords, which were orthographically legal and easy for native speakers to pronounce (*bur*, *ciez*, *daip*, *fot*, *gan*, *huf*, *jom*, *lerg*, *mig*, *nem*, *pes*, *kov*, *sjuk*, *sor*, *talt*, *trin*, *rew*, *vun*, *wijb*, *zas*). The tokens were recorded by a female native Dutch speaker. Assignment of tokens to elements (first, second, or third) was randomized across participants.
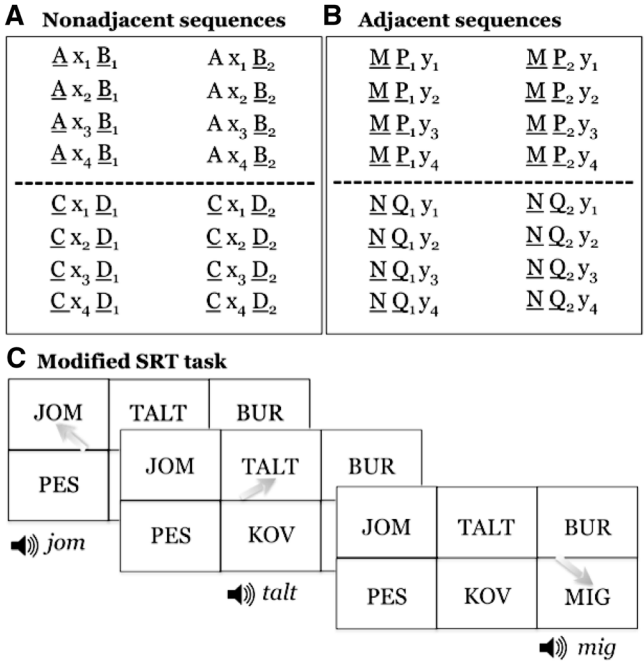
A Nonadjacent sequences   **B** Adjacent sequences

$\underline{A}\,x_1\,\underline{B_1}$	$A\,x_1\,\underline{B_2}$	$\underline{M}\,\underline{P_1}\,y_1$	$\underline{M}\,\underline{P_2}\,y_1$
$\underline{A}\,x_2\,\underline{B_1}$	$A\,x_2\,\underline{B_2}$	$\underline{M}\,\underline{P_1}\,y_2$	$\underline{M}\,\underline{P_2}\,y_2$
$\underline{A}\,x_3\,\underline{B_1}$	$A\,x_3\,\underline{B_2}$	$\underline{M}\,\underline{P_1}\,y_3$	$\underline{M}\,\underline{P_2}\,y_3$
$\underline{A}\,x_4\,\underline{B_1}$	$A\,x_4\,\underline{B_2}$	$\underline{M}\,\underline{P_1}\,y_4$	$\underline{M}\,\underline{P_2}\,y_4$
$\underline{C}\,x_1\,\underline{D_1}$	$\underline{C}\,x_1\,\underline{D_2}$	$\underline{N}\,\underline{Q_1}\,y_1$	$\underline{N}\,\underline{Q_2}\,y_1$
$\underline{C}\,x_2\,\underline{D_1}$	$\underline{C}\,x_2\,\underline{D_2}$	$\underline{N}\,\underline{Q_1}\,y_2$	$\underline{N}\,\underline{Q_2}\,y_2$
$\underline{C}\,x_3\,\underline{D_1}$	$\underline{C}\,x_3\,\underline{D_2}$	$\underline{N}\,\underline{Q_1}\,y_3$	$\underline{N}\,\underline{Q_2}\,y_3$
$\underline{C}\,x_4\,\underline{D_1}$	$\underline{C}\,x_4\,\underline{D_2}$	$\underline{N}\,\underline{Q_1}\,y_4$	$\underline{N}\,\underline{Q_2}\,y_4$

C Modified SRT task

Figure 1 The lists of adjacent and nonadjacent grammatical sequences used in the study are presented in panels A and B. For the SRT task, written distractors were drawn from the other subset of each dependency type (indicated by the dashed line in panels A and B). An example of the modified SRT display is presented in panel C.

Two sets of grammatical sequences were constructed, both with a small adjacent set size (x = 4 exemplars; see Figures 1A and B). There were a total of 16 unique grammatical sequences per dependency set. The nonadjacent sequences consisted of four probabilistic nonadjacent pairings between the first and the third element (A-B_1 and A-B_2; C-D_1 and C-D_2; $p = .50$). As each second element (X_1, X_2, X_3, or X_4) was followed by each third element (B_1, B_2, D_1, or D_2) at equal frequency, the second element did not provide any useful information for selecting the third element. The adjacent sequences consisted of four probabilistic adjacent pairings between the first and the second element (M-P_1 and M-P_2; N-Q_1 and N-Q_2; $p = .50$). As each third element (Y_1, Y_2, Y_3, and Y_4) followed each second element (P_1, P_2, Q_1, or Q_2) at equal frequency, third elements did not provide any useful backward information for selecting the second elements. Adjacent and nonadjacent sequences were randomly intermixed during training.

Modified SRT Task

The modified SRT task was used to obtain online measures of learning (see Misyak & Christiansen, 2010; Misyak et al., 2010a, 2010b). On their computer screen, the participants saw a table with three columns and two rows (see Figure 1C). Each cell showed a nonword. Upon hearing each nonword in a sequence (e.g., "jom-talt-mig"), they simply had to click on the corresponding cells (the targets). The written target for the first element (e.g., JOM) was presented in the first column, the second target (e.g., TALT) in the second column, and the third target (e.g., MIG) in the third column.

Each column of the table included one target and one distractor. The distractors were drawn from the other subset within each dependency type (see Figure 1) and were constrained by element position. The distractors used for the first column were the first elements of a nontarget sequence in the same dependency set (e.g., the written distractor for the target element A was element C and vice versa; for element M, the written distractor was element N and vice versa). Similarly, the distractors for the second and third columns were second and third elements, respectively, of a nontarget sequence in the same dependency set (e.g., the written distractor for the target second element P was element Q and vice versa). The written targets and distractors appeared equally frequently and were counterbalanced for display positions across trials.

A training trial started with the presentation of a fixation cross in the center of the computer screen for 750 milliseconds. Then the visual display was shown until the end of the trial. The first nonword was played 250 milliseconds after the onset of the display. From the starting position of the mouse, which was at the center of the computer display, the participants were asked to make a mouse click inside the rectangular area containing the first target as quickly and accurately as possible. Immediately following the participant's response, the second nonword was played and the participant made a second mouse-click response. The same procedure applied to the third nonword. The trial ended as soon as the participant had made the third mouse click. RTs were measured from the onset of each element's auditory presentation. A different random order of trials was used for each participant. We did not aim to compare RTs across positions (e.g., for the first vs. second target) but were only interested in RT changes for a given position across the trials of the training sessions. In other words, it may be the case that it took participants longer to move the mouse to the first than to the second target, but this was not relevant for the present purposes.

To assess pattern-specific learning, ungrammatical trials were included in addition to grammatical trials. In each session, participants went through four

grammatical blocks (16 grammatical sequences × 4 repetitions = 64 trials per dependency set per block), then one ungrammatical block of 8 nonadjacency and 8 adjacency violation trials, and two further grammatical blocks. Nonadjacent ungrammatical sequences were constructed by switching the third elements across nonadjacent subsets (e.g., Ax<u>D</u>, Cx<u>B</u>). Adjacent ungrammatical items were constructed by switching the second elements across adjacent subsets (e.g., M<u>Q</u>y, N<u>P</u>y). Adjacent and nonadjacent ungrammatical sequences were randomly intermixed in the ungrammatical blocks. Pattern-specific learning was assessed by comparing the RTs in the ungrammatical blocks to the RTs in the grammatical blocks that immediately preceded and followed the ungrammatical ones (e.g., Misyak et al., 2010b). If dependencies among the elements are learned, the RTs should be longer in the ungrammatical block than in the grammatical blocks. RTs for the second and third elements were averaged separately using trials from the appropriate adjacent and nonadjacent sets—henceforth, adjacent and nonadjacent RTs, respectively.

End-of-Training Tests
There were no offline tests after the first or second training sessions. Upon completing the training trials of the third session, the participants were told that the triplets they had heard followed certain patterns and that their knowledge of these regularities would be tested in three short tasks—a nonadjacency prediction task (8 trials) followed by an adjacency prediction task (8 trials) and a grammaticality judgment task (16 trials per dependency set). No feedback was provided for any of the offline tests.

In the prediction tasks, the same SRT display was used as during training, but the critical element was omitted from the spoken stimuli. In the nonadjacency prediction task, the participants made mouse-click responses for the first and second elements based on auditory information, as before, but selected the third element without any auditory information. In other words, they had to predict the third element. For the items in the nonadjacent set, this was possible based on the nature of the first element, whereas the second element was uninformative. In the adjacency prediction task, they made a mouse-click response after hearing the first token, then selected one of the elements in the second position without any auditory information, and finally selected the third elements based on the auditory information.

In the grammaticality judgment task, the participants heard one triplet of spoken elements at a time. Adjacent and nonadjacent trials were randomly intermixed. Half of the triplets within each set followed trained patterns, whereas the remaining half had a violation at the critical second or third element position

(e.g., nonadjacent grammatical AxB vs. ungrammatical AxD, adjacent grammatical MPy vs. ungrammatical MQy). The participants pressed one of two keys (yes or no) to indicate whether or not the triplet followed previously trained patterns. Responses were scored as correct if the appropriate key was pressed.

Similar to Romberg and Saffran (2013), the set of bigrams (e.g., xB, xD) were of equal frequency across our grammatical and ungrammatical nonadjacent sequences. Adjacent probabilities were not informative for distinguishing between grammatical and ungrammatical sequences on nonadjacent test trials. Conversely, nonadjacent probabilities were not informative for the adjacent test trials.

Results

SRT Task

Participants were 98% correct on average ($SD = 1.3$) in making mouse-click responses during training. All correct reaction times were included in subsequent RT analyses. A summary of adjacent and nonadjacent RTs across training blocks and sessions is presented in Figure 2.

The adjacent and nonadjacent RTs decreased considerably across the grammatical blocks, especially within the first session. Collapsing across dependency types, the RT improvement averaged 60 milliseconds ($SD = 55$) within the first session and close to zero in the subsequent sessions. A 2 (Dependency type: adjacent vs. nonadjacent) × 3 (Session: 1 to 3) × 6 (grammatical blocks within each session) within-subject analysis of variance (ANOVA) on logarithmically transformed RTs confirmed that there were significant main effects of session and block (both ps $< .001$) and a significant session × block interaction, which was corrected for violation of the sphericity assumption using the Greenhouse-Geisser formula, $F(6.38, 280.78) = 13.15$, $p < .001$, $\eta_p^2 = .23$. Collapsing across the sessions, the adjacent RTs tended to decrease slightly faster than the nonadjacent RTs over the grammatical blocks (mean RT decrease between the initial and final block $= 25$ milliseconds, $SD = 36$, for adjacent dependencies, compared to 13 milliseconds, $SD = 39$, for nonadjacent dependencies). However, the dependency × session interaction was not significant (Greenhouse-Geisser corrected), $F(4.03, 177.16) = 2.19$, $p = .07$, $\eta_p^2 = .05$. There was no significant main effect of dependency type, $F < 1$, $p = .88$, $\eta_p^2 < .001$, nor was there a significant dependency × session interaction, $F < 1.62$, $p = .20$, $\eta_p^2 = .04$, or a significant three-way interaction, $F < 1$, $p = .76$, $\eta_p^2 = .02$.

If the participants' RTs decreased across blocks and sessions because they had learned the dependencies and could predict the target elements, then removing the regularities in these patterns should lead to disruption of their responses
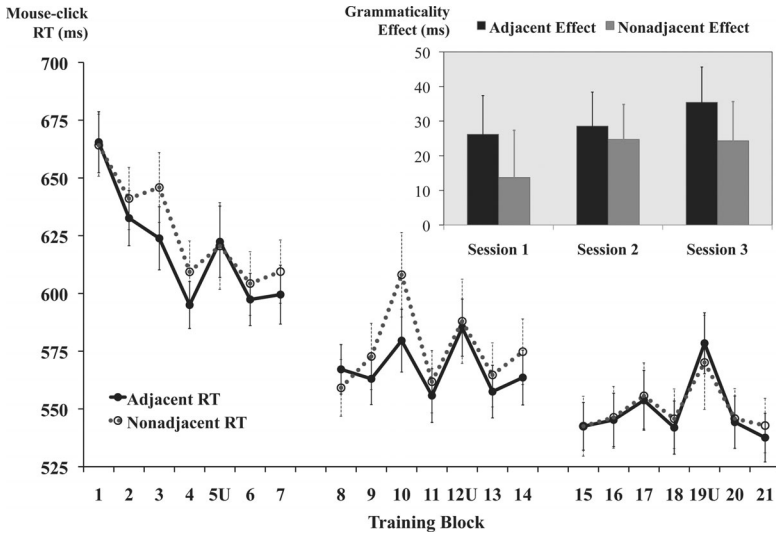
Figure 2 Adjacent and nonadjacent RTs across training blocks (error bars represent standard errors). Each session consisted of seven training blocks. Blocks 5, 12, and 19 were ungrammatical blocks. The bar chart displays the averaged grammaticality effects across dependency types and sessions (grammaticality effect = ungrammatical RT – pooled grammatical RTs, which were based on the grammatical RTs in the two blocks that immediately preceded and followed the ungrammatical block).

in the ungrammatical blocks. Reinstating the regularities should lead to a rebound in response facilitation. To test this prediction, we compared the average RTs in the ungrammatical blocks to the pooled RTs in the grammatical block immediately preceding the ungrammatical block and the grammatical block immediately following the ungrammatical block. A 2 (Dependency type: adjacent vs. nonadjacent) × 3 (Session: 1 to 3) × 2 (Grammaticality: ungrammatical vs. grammatical) within-subject ANOVA on logarithmically transformed RTs showed a significant main effect of session, indicating that the overall mean RTs decreased with training (Greenhouse-Geisser corrected), $F(1.66, 73.19) = 22.44, p < .001, \eta_p^2 = .34$. Crucially, the ANOVA confirmed that the main effect of grammaticality was significant, $F(1, 44) = 39.39, p < .001, \eta_p^2 = .47$. Collapsing across dependency type and sessions, grammatical RTs were faster than ungrammatical RTs by 25 milliseconds on average ($SD = 33$). However, there was no main effect of dependency type, $F < 1, p = .49, \eta_p^2 = .01$, nor was there a significant dependency × grammaticality interaction, $F(1, 44) = 2.92, p = .10, \eta_p^2 = .06$, or a significant three-way interaction, $F < 1, p = .79, \eta_p^2 = .005$.

Table 1 Summary of the descriptive statistics and the fixed effects in the logistic regressions for the offline data

A. Descriptive statistics for prediction and grammaticality judgment data

Measure	Dependency type	Accuracy (*SD*)	Hit (*SD*)	False alarm (*SD*)	d' (*SD*)
Prediction	All trials	61% (18)	–	–	–
	Adjacent	64% (23)	–	–	–
	Nonadjacent	58% (23)	–	–	–
Grammaticality judgment	All trials	58% (10)	–	–	–
	Adjacent	63% (14)	.69 (.19)	.43 (.20)	.88 (1.05)
	Nonadjacent	54% (13)	.66 (.19)	.58 (.18)	.27 (.83)

B. Fixed effects for the prediction data ($N = 720$)

Model	Predictor	Coefficient	*SE*	Wald Z	*p*
Intercept-only	Intercept	.48	.12	3.94	<.001
Dependency as predictor	Intercept	.33	.14	2.29	.02
	Dependency = adjacent	.31	.16	1.94	.05

C. Fixed effects for the grammaticality judgment data ($N = 1440$)

Model	Predictor	Coefficient	*SE*	Wald Z	*p*
Intercept-only	Intercept	.34	.06	5.40	<.001
Dependency as predictor	Intercept	.16	.08	1.98	<.05
	Dependency = adjacent	.37	.11	3.40	<.001

Offline Results

A summary of the offline results can be found in Table 1. The participants averaged 61% correct (*SD* = 18) on the prediction task and 58% correct (*SD* = 10) on the grammaticality judgment task. To test whether the overall performance differed from chance (50%) for each task, we carried out a mixed-effects logistic regression using the *lme4* package in the statistical software R (Bates, Maechler, & Bolker, 2011; Jaeger, 2008). Raw accuracy data were entered as the dependent variable, subjects as a random effect, and only the intercept included in the fixed effects. Consistent with the SRT results, the regression showed that the participants performed significantly above chance on both the prediction task, $b = .48$, $z = 3.94$, $p < .001$, and the grammaticality judgment task, $b = .34$, $z = 5.40$, $p < .001$.

In contrast to the SRT results, however, the participants showed lower accuracy on the nonadjacent than adjacent dependencies in the offline tasks. On average, the participants scored 58% correct (*SD* = 23) on nonadjacent and 64% correct (*SD* = 23) on adjacent trials in the prediction task, and 54% correct (*SD* = 13) on nonadjacent and 63% correct (*SD* = 14) on adjacent trials in the grammaticality judgment task. A histogram showing the distribution of

A Prediction   ■ Adjacent   ■ Nonadjacent
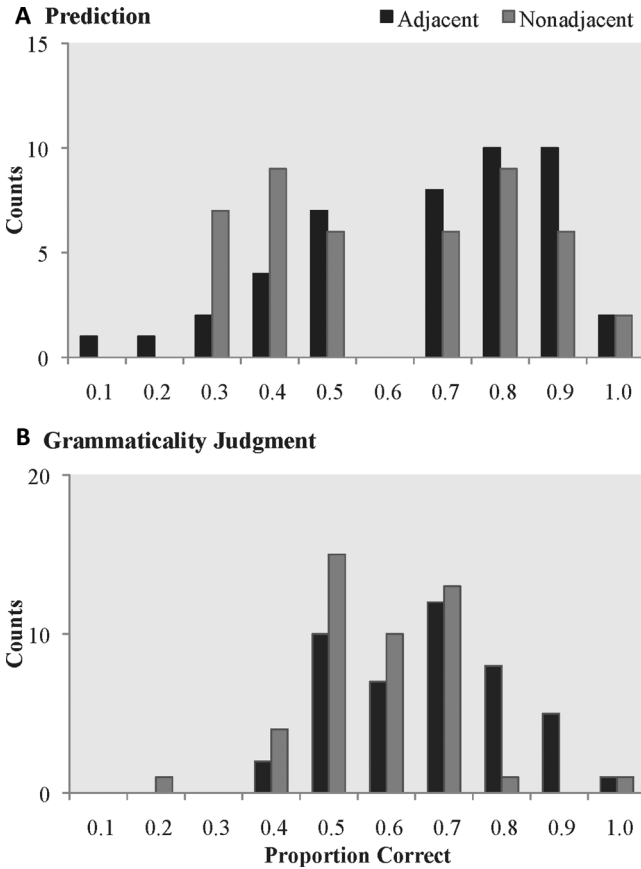
B Grammaticality Judgment

Figure 3 Histograms of the proportion correct (bin width = 0.1) for adjacent and nonadjacent trials in (A) the prediction task and (B) the grammaticality judgment task.

the proportions of correct values by the dependency type in each task can be found in Figure 3.

To examine whether offline performance differed as a function of the type of dependencies tested, we performed another mixed-effects logistic regression with dependency type added as a fixed effect. As summarized in Table 1, dependency type had a significant effect on offline performance. The odds of making a correct response in the prediction task were 1.36 times higher for adjacent than nonadjacent dependencies ($b = .31$). Similarly, the odds of making a correct grammaticality judgment were 1.44 times higher for adjacent than nonadjacent trials ($b = .37$). The effect of dependency type

approached significance in the prediction task, $z = 1.94, p = .052$, and was highly significant in the grammaticality judgment task, $z = 3.40, p < .001$. The significant intercepts in the models confirmed that performance on nonadjacent dependencies was significantly above chance in both tasks (both ps $< .05$).

Finally, an analysis of the grammaticality judgment data using signal detection methods (e.g., Stanislaw & Todorov, 1999) yielded a converging pattern of better sensitivity favoring adjacent dependencies in offline performance (see Table 1). The analysis showed comparable average hit rates across dependency types ($M = .69$ and $.66$, respectively), but a much higher false alarm rate for nonadjacent than adjacent dependencies ($M_{adjacent} = .43, M_{nonadjacent} = .58$). A t test confirmed that the resulting d-primes were significantly higher for the adjacent than nonadjacent condition ($M_{effect} = .61, SD = 1.29$), $t(44) = 3.16, p = .003$, although both measures were significantly higher than 0 (both ps $< .05$), indicating again that participants were able to discriminate between the grammatical and ungrammatical sequences for both dependency types, but that their discrimination performance was better for adjacent than nonadjacent dependencies.

Discussion

Participants in the present study heard nonword sequences featuring adjacent dependencies and sequences featuring nonadjacent dependencies with a small intervening set size (four elements). Both the online data (based on RTs in the SRT task) and the offline tests (prediction and grammaticality judgment) showed that the participants learned the adjacent dependencies. Successful learning of the adjacent dependencies is consistent with the results of many earlier studies (e.g., Saffran et al., 1996; Teinonen et al., 2009).

More importantly, the participants in our study also acquired nonadjacent dependencies, which have often been found much harder to learn (see Aslin & Newport, 2009, for a review). Earlier studies where successful nonadjacent learning was observed used preexposure of the nonadjacent pairs (e.g., Lany et al., 2007; Lany & Gómez, 2008), perceptual cues to facilitate their grouping (e.g., Creel et al., 2004; van den Bos et al., 2012), or targeted larger intervening set sizes than used here (e.g., Gómez, 2002; Romberg & Saffran, 2013; van den Bos et al., 2012). Although the finding showing that nonadjacent learning is possible under low intervening variability without added cues does not constitute a novel observation (e.g., De Diego Balaguer et al., 2007), our results extended previous findings, as we observed learning of probabilistic nonadjacent dependencies under low intervening variability while earlier studies used only deterministic ones. Together the findings suggest

that, although nonadjacent learning can be facilitated by the aforementioned conditions, none of these conditions are necessary for learning to occur.

Research on early language development has shown that important elements, such as individual verbs in verb-centered grammatical structures, tend to show low variability in early stages of language acquisition (e.g., Goldberg, Casenhiser, & Sethuraman, 2004; Ninio, 1999; see also N. C. Ellis & Ferreira-Junior, 2009; Wulff, N. C. Ellis, Romer, Bardovi-Harlig, & Leblanc, 2009, for studies on second language acquisition). Ninio (1999) focused on the development of verb-object (VO) and subject-verb-object (SVO) patterns in younger children (aged 1;01 to 2;01) and found that children typically started with one to two path-breaking verbs. The children continued to use these verbs for a relatively long period of time before they added more verbs to the patterns. Similarly, Goldberg et al. (2004) performed corpus analyses of children's and mothers' speech and found that most of the instances in which particular patterns were used involved a limited number of verbs—a highly frequent verb together with several less frequent ones. These and related studies (e.g., Boyd & Goldberg, 2009; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008; Wulff et al., 2009) have argued that the initial learning of input patterns can be robust with, and may be even facilitated by, low-variance input. While previous studies on the learning of nonadjacent dependencies has shown that nonadjacent learning is facilitated by more, rather than less, variable input (e.g., Gómez, 2002), our finding is relevant to the language research discussed here as it shows that statistical learning of nonadjacent relations can be robust in the face of low (intervening) variability.

In our study, the probabilities of the dependency patterns in the adjacent and nonadjacent sets were matched. In accordance with these matched probabilities, the online results indicated that the nonadjacent dependencies were learned as well as the adjacent ones. Neither the online nor offline data showed evidence for a mutually exclusive pattern of adjacent versus nonadjacent learning. These findings are therefore inconsistent with the view that learners focus by default on adjacent dependencies, which would hinder nonadjacent learning (Gómez, 2002; see also Pacton & Perruchet, 2008). They are, however, in line with a view of associative learning that postulates the formation of "remote associations" between various nonadjacent elements in a sequence along with adjacent associations between contiguous elements, as put by Ebbinghaus (1885/1964) over a century ago:

> [T]he associative threads, which hold together a remembered series, are
> spun not merely between each member and its immediate successor, but

> beyond intervening members to every member which stands to it in any
> close temporal relation. (p. 94)

In accord with this manifold associative learning view (Ebbinghaus, 1885/1964; see Slamecka, 1985, for a review; for discussions, see Capaldi, 1985; Dallett, 1965; N. C. Ellis, 1970; Hakes & Young, 1966; Slamecka, 1964, 1965), the present study, combined with Romberg and Saffran's (2013) findings, shows that adjacent and nonadjacent learning can occur together when learners are exposed to strings with adjacent and nonadjacent patterns. This concurrent learning is possible even under low intervening variability, as indicated by our results.

Moreover, Romberg and Saffran (2013) found that adjacent and nonadjacent learning can occur together when probabilistic adjacent and deterministic nonadjacent dependencies are embedded within the same strings. Our results further show that adjacent and nonadjacent learning can proceed in parallel when probabilistic adjacent and probabilistic nonadjacent dependencies are present in different strings. This suggests that statistical learning can exploit transitional probabilities across multiple sets of dependency patterns concurrently, for each set that contains predictive information, be it adjacent or nonadjacent dependencies. Previously, van den Bos et al. (2012) found no learning of probabilistic nonadjacent dependencies under limited exposure (one session of approximately 20 minutes) in the absence of grouping cues. The successful learning found here highlights the importance of providing extended exposure to the patterns. In our study, the forward conditional probabilities were matched at $p = .50$ for both adjacent and nonadjacent dependencies. In addition, however, the backward conditional probabilities were 1.0 for both types of dependencies.[1] Prior studies have shown that learners are also sensitive to backward conditional probabilities (e.g., Pelucchi, Hay, & Saffran, 2009; Perruchet & Desaulty, 2008), and hence it is possible that the backward statistics might play a role in the current results (though whatever effect this may have, this effect will be constant across both types of dependencies). Future studies should look into the possible contributions of these sources of predictability to the learning of adjacent and nonadjacent dependencies.

Our study included both online and offline measures of learning. In neither of these measures was there evidence for an all-or-none trade-off between adjacent and nonadjacent learning. Nevertheless, there was a difference in the offline results favoring adjacent dependencies, which was absent in the online SRT measure. What might account for this pattern? One possibility is that our SRT measure failed to capture the adjacency advantage in the online learning

phase. In this study, the ungrammatical trials were implemented toward the end of each training session. Visual inspection of the SRT curves suggests some divergence between the adjacent and nonadjacent curves earlier, such as at the third block of the first session, although the lack of ungrammatical RT control did not allow for a measure of dependency-specific learning to be derived and tested. Perhaps adjacent dependencies can be learned faster than nonadjacent dependencies thanks to their greater temporal proximity. However, such an advantage in the speed of learning may be only temporary, especially for very short sequences like those used here. If this possibility is correct, we may be able to detect online differences favoring adjacent dependencies when dependency-specific learning measures are administered earlier in training.

Alternatively, the adjacency advantage might arise specifically in the offline tests. The same might apply for earlier studies suggesting an adjacency bias in offline measures of learning (e.g., Gómez, 2002). For sequences as short as ours, it is possible that an adjacency advantage may be negligible in online measures. During the offline tests, however, participants had to make an explicit choice for which they had to translate their intuitions about the sequences into overt responses. In this task, the adjacent pairs may have been more salient or easier to retrieve from working memory than nonadjacent pairs. This interpretation is compatible with the view that adjacent associations may be more amenable to direct observations than remote associations (Ebbinghaus, 1885/1964; Slamecka, 1985). However, Ebbinghaus postulated that adjacent associations may be the *only* ones observed in "conscious mental life" (p. 94). This claim is not entirely consistent with our results. Following conventional interpretation of the signal detection sensitivity measure, the above-zero d-prime for nonadjacent dependencies seen here suggests that nonadjacent dependencies *can* influence performance on direct tests, though to a lesser extent compared to adjacent dependencies. Clearly, more research is necessary to shed further light on the adjacency advantage found in offline performance (and the lack thereof in online data). With respect to our main question—whether adjacent and nonadjacent sequences from separate sets of materials can both be learned in the same exposure window in spite of low intervening variability—both types of measures yielded converging positive evidence.

Implications and Conclusion

An emerging body of empirical studies has begun to link statistical learning to natural language acquisition and processing (see Arciuli & Torkildsen, 2012, for a review). However, these studies have tended to focus on first language—linking statistical learning, for instance, to grammar acquisition (Kidd, 2012),

vocabulary development (E. M. Ellis, Gonzales, & Deák, 2014; Shafto, Conway, Field, & Houston, 2012), and online comprehension in first language (Conway, Bauernschmidt, Huang, & Pisoni, 2010; Misyak & Christiansen, 2012; Misyak et al., 2010a, 2010b). The robust learning of a relatively complex set of dependency patterns shown by our young adult participants suggests that statistical learning may play a role in adult second language learning (see also Onnis, 2012). Future investigations should look beyond first and child language acquisition, and begin to examine the role of statistical learning in adult and second language acquisition. In investigating these links, our study suggests that investigators should carefully plan duration of exposure so as to ensure sufficient exposure to trained patterns. It is also advisable that online measures of learning be used in addition to offline ones. Insofar as the learning of complex patterns is concerned, the sole use of offline measures risks providing an incomplete assessment of learning (e.g., Jiménez et al., 1996; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Tokowicz & MacWhinney, 2005; see also Norris & Ortega, 2000, for a review indicating biases in the measures used to assess learning in second language acquisition research). Inclusion of online measures is all the more pressing, as research begins to move toward exploring the role of statistical learning in the acquisition of such complex systems as natural language.

To conclude, our results show that, under suitable learning conditions, adult learners acquire nonadjacent dependencies readily along with adjacent ones. There was, in our data, no evidence that one type of learning occurred at the expense of the other type of learning. Our results indicate that statistical learning is more powerful than previously thought, which supports the hypothesis that statistical learning may play an important role in the acquisition of long-distance dependencies in natural language.

<div align="right">Final revised version accepted 8 September 2014</div>

Note
1 We thank an anonymous reviewer for pointing this out.

References
Arciuli, J., & Torkildsen, J. (2012). Frontiers: Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data. *Frontiers in Language Sciences*, *3*, 1–9. doi:10.3389/fpsyg.2012.00324

Aslin, R. N., & Newport, E. L. (2009). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language* (pp. 15–29). New York: Psychology Press.

Bates, D., Maechler, M., & Bolker, B. (2011). Lme4: Linear mixed-effects models using S4 classes. R version 2.14.1 (2011-12-22). Retrieved from http://CRAN.R-project.org/package=lme4

Boyd, J. K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *The Modern Language Journal*, *93*, 418–429. doi:10.1111/j.1540-4781.2009.00899.x

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*, 127–132. doi:10.1016/j.cognition.2011.06.010

Capaldi, E. J. (1985). Anticipation and remote associations: A configural approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 444–449. doi:10.1037/0278-7393.11.3.444

Cherry, K. E., & Stadler, M. E. (1995). Implicit learning of a nonverbal sequence in younger and older adults. *Psychology and Aging*, *10*, 379–394. doi:10.1037/0882-7974.10.3.379

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356–371. doi:10.1016/j.cognition.2009.10.009

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1119–1130. doi:10.1037/0278-7393.30.5.1119

Dallett, K. M. (1965). In defense of remote associations. *Psychological Review*, *72*, 164–168. doi:10.1037/h0021774

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A. C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE*, *2*(11), e1175. doi:10.1371/journal.pone.0001175

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin and Review*, *4*, 3–23. doi:10.3758/BF03210769

Durrant, S. J., Taylor, C., Cairney, S. & Lewis, P.A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*, 1322–1331.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover. (Original work published 1885; translated 1913).

Ellis, E. M., Gonzalez, M. R., & Deák, G. O. (2014). Visual prediction in infancy: What is the association with later vocabulary? *Language Learning and Development*, *10*, 36–50. doi:10.1080/15475441.2013.799988

Ellis, N. C. (1970). Serial task structure and the doctrine of remote associations. *Psychonomic Science*, *21*, 121–122. doi:10.3758/BF03335796

Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, *93*, 370–385. doi:10.1111/j.1540-4781.2009.00896.x

Gaillard, V., Destrebecqz, A., Michiels, S., & Cleeremans, A. (2009). Effects of age and practice in sequence learning: A graded account of ageing, learning, and control. *European Journal of Cognitive Psychology*, *21*, 255–282. doi:10.1080/09541440802257423

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin & Review*, *16*, 486–490. doi:10.3758/PBR.16.3.486

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, *15*, 289–316. doi:10.1515/cogl.2004.011

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436. doi:10.1111/1467-9280.00476

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*, 183–206. doi:10.1207/s15327078in0702_4

Hakes, D. T., & Young, R. K. (1966). On remote associations and the interpretation of derived-list experiments. *Psychological Review*, *73*, 248–251. doi:10.1037/h0023220

Howard, D. V., & Howard, J. H. (1989). Age differences in learning serial patterns: Direct versus indirect measures. *Psychology and Aging*, *4*, 357–364. doi:10.1037/0882-7974.4.3.357

Howard, D. V., & Howard, J. H. (1992). Adult age differences in the rate of learning serial patterns: Evidence from direct and indirect tests. *Psychology and Aging*, *7*, 232–241. doi:10.1037/0882-7974.7.2.232

Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology*, *5*, 175. doi:10.3389/fpsyg.2014.00175

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. doi:10.1016/j.jml.2007.11.007

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 948–969. doi:10.1037/0278-7393.22.4.948

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*, 171–184. doi:10.1037/a0025405

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence of a domain general learning mechanism. *Cognition*, *83*, B35–B42. doi:10.1016/S0010-0277(02)00004-5

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*, 1247–1252. doi:10.1111/j.1467-9280.2008.02233.x

Lany, J., Gómez, R. L., & Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, *31*, 481–507. doi:10.1080/15326900701326584

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*, 628–634. doi:10.1111/j.1467-7687.2008.00707.x

Misyak, J. B., & Christiansen, M. H. (2010). When "more" in statistical learning means "less" in language: Individual differences in predictive processing of adjacent dependencies. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2686–2691). Austin, TX: Cognitive Science Society.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*, 302–331. doi:10.1111/j.1467-9922.2010.00626.x

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010a). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, *2*, 138–153. doi:10.1111/j.1756-8765.2009.01072.x

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010b). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*, 1–9. doi:10.3389/fpsyg.2010.00031.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*, 933–947. doi:10.1162/jocn_a_00119

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127–162. doi:10.1016/S0010-0285(03)00128-2

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language*, *26*, 619–653.

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

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528. doi:10.1111/0023-8333.00136

Onnis, L. (2012). The potential contribution of statistical learning to second language acquisition. In J. Williams & P. Trebuchat (Eds.), *Statistical learning and language acquisition* (pp. 203–235). Berlin, Germany: Mouton de Gruyter.

Onnis, L., Christiansen, M., Chater, N., & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 886–891). Mahwah, NJ: Erlbaum.

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*, 80–96. doi:10.1037/0278-7393.34.1.80

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*, 244–247. doi:10.1016/j.cognition.2009.07.011

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory and Cognition*, *36*, 1299–1305. doi:10.3758/MC.36.7.1299

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*, 1290–1320. doi:10.1111/cogs.12050

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928. doi:10.1126/science.274.5294.1926

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, *275*, 1599–1603.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*, 247–271. doi:10.1111/j.1532-7078.2011.00085.x

Slamecka, N. J. (1964). An inquiry into the doctrine of remote associations. *Psychological Review*, *71*, 61–76. doi:10.1037/h0048854

Slamecka, N. J. (1965). In defense of a new approach to old phenomena. *Psychological Review*, *72*, 242–246. doi:10.1037/h0021799

Slamecka, N. J. (1985). Ebbinghaus: Some associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 414–435. doi:10.1037/0278-7393.11.3.414

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149. doi:10.3758/BF03207704

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21. doi:10.1186/1471-2202-10-21

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*, 173–204. doi:10.1017/S0272263105050102

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*, 507–520. doi:10.1016/j.jml.2012.07.008

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2011). Simultaneous online tracking of adjacent and non-adjacent dependencies in statistical learning. In L. Carlson, C. Holscher, & T. Shipley (Eds.), Proceedings of the 33rd annual

conference of the cognitive science society (pp. 964–969). Austin, TX: Cognitive Science Society.

Willingham, D. B., Greeley, T., & Bardone, A. M. (1993). Dissociation in a serial response time task using a recognition measure: Comment on Perruchet and Amorim (1992). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1424–1430. doi:10.1037/0278-7393.19.6.1424

Wulff, S., Ellis, N. C., Römer, U., Bardovi-Harlig, K., & Leblanc, C. J. (2009). The acquisition of tense–aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal*, *93*, 354–369. doi:10.1111/j.1540-4781.2009.00895.x