



# Top-down grouping affects adjacent dependency learning

Felix Hao Wang<sup>1</sup> · Jason D. Zevin<sup>2,3</sup> · John C. Trueswell<sup>4</sup> · Toben H. Mintz<sup>2,3</sup>

Published online: 15 June 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

A large body of research has demonstrated that humans attend to adjacent co-occurrence statistics when processing sequential information, and bottom-up prosodic information can influence learning. In this study, we investigated how top-down grouping cues can influence statistical learning. Specifically, we presented English sentences that were structurally equivalent to each other, which induced top-down expectations of grouping in the artificial language sequences that immediately followed. We show that adjacent dependencies in the artificial language are learnable when these entrained boundaries bracket the adjacent dependencies into the same sub-sequence, but are not learnable when the elements cross an induced boundary, even though that boundary is not present in the bottom-up sensory input. We argue that when there is top-down bracketing information in the learning sequence, statistical learning takes place for elements bracketed within sub-sequences rather than all the elements in the continuous sequence. This limits the amount of linguistic computations that need to be performed, providing a domain over which statistical learning can operate.

**Keywords** Dependency learning · Rhythm · Statistical learning

How do learners perform word segmentation to break into language when the linguistic input they have access to is a continuous language stream? One prominent line of research within cognitive science has aimed to understand the role that statistical learning, and, in particular, statistical distributional analysis, plays in discovering underlying structure from serial input, and the role such processes may play in language acquisition. It has been shown, for example, that sequential co-occurrence statistics in speech and other auditory streams are computed by human infants (Aslin, Saffran & Newport, 1998; Saffran, Aslin, & Newport, 1996; Johnson & Jusczyk, 2001) and adults (Mirman, Magnuson, Graf Estes, & Dixon, 2008; Romberg & Saffran, 2013). These findings

led to proposals that humans are capable of using sequential co-occurrence statistics to segment continuous speech into words, using only bottom-up statistical information (Aslin et al., 1998; Estes, Evans, Alibali, & Saffran, 2007; Saffran et al., 1996). Other work has demonstrated how bottom-up cues such as prosody interact with statistical learning (Johnson & Jusczyk, 2001; Johnson & Seidl, 2009; Morgan, Meier, & Newport, 1987, 1989; Shukla, Nespore, & Mehler, 2007). Many of these studies find that bottom-up prosodic cues serve to bracket the incoming sequence. For example, Christophe, Peperkamp, Pallier, Block, and Mehler (2004) showed that there is a delay to lexical access when the syllables in the word crosses a phonological boundary. Palmer, Hudson, White, and Mattys (2019) showed the statistical learning can benefit from having a known word in the statistical learning stream, presumably because the known word provides bracketing of the continuous stream into many sub-sequences that are shorter and easier to process. Mattys, White, and Melhorn (2005) developed a hierarchical framework in which different cues for segmentation, where cues are hierarchically integrated, with descending weights allocated to lexical, segmental, and prosodic cues.

Importantly, the physical stimuli in those experiments that manipulated bottom-up cues (e.g., the presence or absence of a phrase boundary between syllables of the word) were different between different conditions. Indeed, it could not have

✉ Felix Hao Wang  
hao.wang@unlv.edu

<sup>1</sup> Department of Psychology, University of Nevada, Las Vegas, Las Vegas, NV 89154, USA

<sup>2</sup> Department of Psychology, University of Southern California, Los Angeles, CA 90089, USA

<sup>3</sup> Department of Linguistics, University of Southern California, Los Angeles, CA 90089, USA

<sup>4</sup> Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

been any other way, since bottom-up cues are by their nature linked to properties of the physical signal. In contrast, in this paper, we study how one physical stimulus is parsed differently as the result of differences in top-down expectations, and how that results in different outcomes of statistical learning. We know of only one other study of the effect of top-down effect on statistical learning of adjacent dependencies (Lew-Williams, Pelucchi, & Saffran, 2011) where infants were exposed to bisyllabic or trisyllabic words presented in isolation. When they subsequently heard a continuous sequence made of either bisyllabic or trisyllabic words, they only learned when the word length in the continuous sequence matched the length of the words they first heard in isolation. Thus, the expectation of word length had a powerful effect on subsequent learning.

The current experiment investigates a different aspect of statistical learning in the context of top-down expectations. Specifically, we ask whether different top-down expectations can lead the learner to parse identical physical stimuli differently, given subtly different expectations that are influenced by prior linguistic material. We used the entrainment paradigm developed by Ding, Melloni, Zhang, Tian, and Poeppel (2016a) that has been shown to produce periodic activity in listeners' brains reflecting the abstract phrase structure of a sequence of sentences. Previously, we have shown that nonadjacent dependency learning is facilitated or disrupted depending on, respectively, whether the to-be-learned dependencies are represented within a bracketed sequence or crossing a bracketing boundary (Wang, Zevin, & Mintz, 2017). Here, we test whether the learning of adjacent dependencies, which have been found to be much more robust in the past literature, can also be preserved or disrupted under these conditions. If bracketing from a top-down source precedes the computation of transitional probability, we predict that listeners learn conditional probabilities only when the dependencies are presented in phase with the entraining rhythm (i.e., when the elements making up the adjacent dependencies belong to the same sub-sequence), and not when the elements cross an induced boundary.

## Method

**Subjects** Twenty-four undergraduate students at the University of Southern California were recruited for each of two conditions (*in-phase condition/out-of-phase condition*, described in detail below) from the psychology department subject pool. The study was approved by the Institutional Review Board at the University of Southern California.

**Stimuli** We recorded speech from a native English speaker and digitized the recording at a rate of 44.1 kHz. We recorded two types of words: English words and novel words. For English

words, we recorded five names (Brian<sup>1</sup>, John, Kate, Nate, Clair), five monosyllabic verbs in third-person singular form (turns, keeps, puts, lets, has), five pronouns (these, those, this, that, it), and five particles (down, on, up, off, in). For the artificial language, we used 12 novel words to construct four-word sequences. The four-word sequences had fixed sets of words, where each set appeared in a fixed position: (pel, tink, blit) at Position 1, (swech, voy, rud) at Position 2, (tiv, ghire, jub) at Position 3, and (dap, wesh, tood) at Position 4. The English structure (e.g., John turns these down) was chosen specifically to avoid a phrase-level dependency between the third and the fourth position. The words were recorded in a random order and spliced from the recording, each word by itself lasting between 300 ms and 737 ms. The words were then shortened to 250 ms using the *lengthen* function in Praat (Boersma, 2001). An additional 83ms of silence was added to the end of each word to increase intelligibility, such that when words were concatenated in a continuous stream, they occurred at 3 Hz.

It is worth noting that while the 3-Hz rate is in the range of the syllable rate in natural speech (which can be up to 8 Hz; see Ding, Patel, et al., 2016b), our stimuli might seem fast on the first hearing. The main cause of this initial impression is likely the fact that we intentionally removed prosody from the speech by compressing the syllables to achieve uniform durations. This was a necessary part of the experimental design, intended to eliminate prosodic cues. Informal discussion with participants and other individuals who listened to the materials assured us that listeners get used to this stimuli presentation rate rapidly after a brief adaptation period.

**Apparatus** The experiment ran on a computer using MATLAB 2013B and the Psychophysics Toolbox (Brainard, 1997). Auditory stimuli were presented through headphones, and subjects entered their responses through the keyboard. Instructions were delivered visually on the computer monitor.

**Design and procedure** The experiment was run in three blocks, each composed of a training phase followed by a testing phase. Each condition included two counterbalancing orders, with half of the participants run in each, described below.

**Training phase** Each artificial language sentence was a concatenation of four nonce words, one each from choices of three for each position, as specified in the Stimuli section. Schematically, the structure was  $XYA_iB_j$ , where each letter represents a positional word class, and the specific choice of word in the third position ( $A_i$ ) perfectly determined the

<sup>1</sup> We regret the oversight that Brian is not a monosyllabic word. When we controlled for this fact elsewhere (Wang et al., 2017), this did not make any difference.

- a** Kate turns these off Clair turns that down Brian turns these down  
 John puts that on Nate has it up Claire lets this in John keeps those  
on *pel swech tiv wesh tink voy ghire tood blit rud tiv wesh* Nate puts  
those on Brian has this up . . .
- b** turns these off Clair turns that down Brian turns these down  
 John puts that on Nate has it up Claire lets this in John keeps those  
on Kate *pel swech tiv wesh tink voy ghire tood blit rud tiv wesh* puts  
those on Brian has this up . . .

**Fig. 1** Design of language materials in the training phase. English materials have sentence structures that create perceptual groupings at the sentence level and a rhythmically induced boundary between sentences. English sentences are colored black, and every other sentence is underlined to show grouping. The artificial language sequences are bracketed by color, where the boundaries are shown given the phase of the English sentences. Dependencies in the artificial language are indicated with arrows beneath the text. In the *in-phase*

condition (a), the sub-sequences contain the dependencies to be learned. In particular, the third word always predicted the fourth word, each with a conditional probability of 1.0 (e.g., the probability of hearing “tood” given “jub” was 1.0). In the *out-of-phase* condition (b), the sequence of nonsense words interrupts a sentence, so that the to-be-learned material is not bracketed by the induced rhythm, even though the conditional probability between the third and the fourth word is also 1.0

specific word in the fourth position ( $B_i$ ), for all  $i = 1, 2, 3$ . That is,  $A_1$  was always followed by  $B_1$ ,  $A_2$  was always followed by  $B_2$ , and  $A_3$  was always followed by  $B_3$ . In terms of conditional probability, this means that  $p(B_i | A_i)$  equaled 1.0. Given that there were three different words that could appear in each nondependent position (from classes X, Y, and A), there were 27 possible different quadruplet artificial sentences.

All English sentences had the structure [Name Verb Pronoun Particle] and were created by concatenating a randomly selected name, verb, pronoun, and particle, in that order. As such, each sentence consisted of four monosyllabic words (with the exception of sentences containing the word “Brian”), and lasted 1.33 seconds. Since words were randomly selected and constrained only by position, there were no word-level statistical dependencies within the English sentences themselves.

The training stream was made by concatenating English and artificial language sentences in the following way. We concatenated together five, six, or seven English sentences. After the English sentences, three artificial language sequences were concatenated, followed by another set of English sentences, followed by three more artificial language sequences, and so on (see Fig. 1). There were no additional pauses between English words, between novel words of the artificial language, or at the boundaries between English words or novel words.

For the *in-phase* condition, the English sentences were concatenated starting with the first word in the first sentence, and ending with the last word of the last sentence, so that the sequences of nonsense words began directly after the completion of a sentence cycle and the onset of a new cycle. Thus, the to-be-learned sequences were in phase with the rhythm induced by the sentence boundaries of the English stimuli. We

predicted that subjects would continue to segment the artificial language sequence following that rhythm, which we predicted would facilitate learning of the dependencies in these sequences. In contrast, in the *out-of-phase* condition, we manipulated the English sentences so that the to-be-learned sequences were out of phase with the rhythm produced by the sentence boundaries of the English stimuli. Specifically, for each set of five, six, or seven English sentences, we moved the first word of the entire sequence, a name, and placed at the end of the entire sequence. Thus, the first three words of the set were essential a verb phrase fragment (e.g., *turns these off*), after which there were several four-word-long complete sentences (at least four of them). Hence, we expected subjects to become entrained to the four-word cycle as they started to parse these sentences, as in the *in-phase* condition, but here the cycle is shifted.<sup>2</sup> The artificial word sequences were identical to those in the *in-phase* condition. Thus, in contrast to the *in-phase* condition, we expected subjects to group the first three words of the artificial sequence with the last word of the English sequence, and to continue to segment the artificial stream in that shifted phase (see Fig. 1). Critically, the to-be-learned artificial language dependencies were placed where the expected sentence boundaries fell. We predicted these virtual boundaries would interfere with learning of the dependencies in these sequences.

A total of 432 artificial language sentences and 858 English sentences were divided equally between three blocks, and sequenced in the manner just described.

<sup>2</sup> Although the English sequences in the out-of-phase condition do start with a sentence fragment, the subsequent English sentences are easily parsed, and the four-word rhythm rapidly becomes apparent. The sample auditory stimuli can be found at <https://osf.io/x9msp/>.

For each of the *in-phase* and *out-of-phase* conditions, we created two counterbalancing languages by taking three pairs of AB for the three frames in one training language ( $A_1B_1$ ,  $A_2B_2$ ,  $A_3B_3$ ), and three different pairs for the other training language ( $A_1B_2$ ,  $A_2B_3$ ,  $A_3B_1$ ). This design allowed us to use the same set of test items for the two counterbalancing conditions, where the test set included frames from both training languages. Each block of training material was approximately 9 minutes long.

**Test phase** Immediately after each training block, we displayed instructions for the test phase. The instructions indicated that participants would hear a number of sound sequences and make judgments about the sequences. There were 18 test trials per block, each testing a single item. Half of the items were grammatical sentences of the artificial language, and the other half were ungrammatical. The ungrammatical sentences had AB sequences that were taken from the counterbalanced training condition and had thus not been heard in the training phase. The presentation sequence of test trials was randomized for each participant.

Participants initiated each test trial. At the start of each trial, an artificial language sequence was played, and participants were asked to answer the following question: “Do you think that you heard this sequence in the previous section?” Their responses were marked on a scale containing five response items: *definitely*, *maybe*, *not sure*, *maybe not*, and *definitely not*. Once the participants clicked on any of the choices, the screen went blank and the trial ended. After a 1-second inter-trial interval, the next trial began.

## Results

We coded the scale of *definitely*, *maybe*, *not sure*, *maybe not*, and *definitely not* into the numeric values 4, 3, 2, 1, and 0, respectively.<sup>3</sup> For the *in-phase* condition, the mean rating of grammatical sentences (3.074) was higher (i.e., greater endorsement) than the mean rating of ungrammatical sentences (2.676). Note that the fact that the mean rating of ungrammatical sentences is above zero is not surprising, since the words in the ungrammatical sentences are all familiar words; it is only the co-occurrence between two words that is novel. We should not expect participants to completely reject the ungrammatical sentences based on the ungrammatical dependencies. What matters is the prediction that grammatical sentences should be rated more familiar than the ungrammatical sentences.

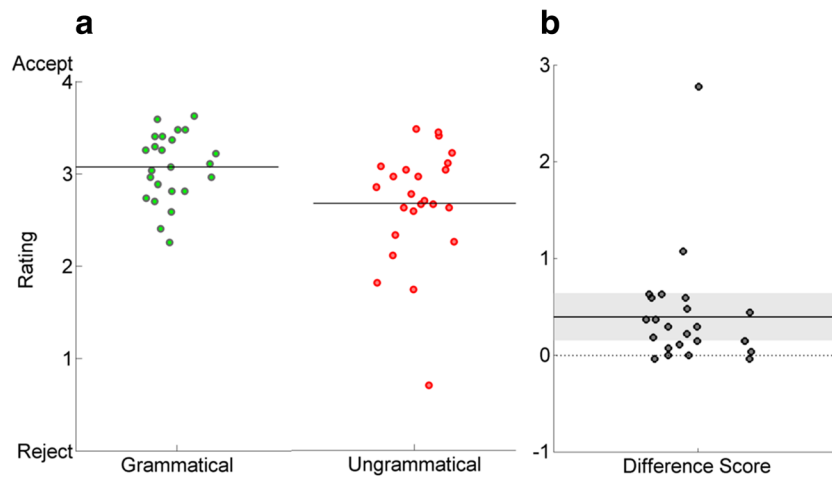
To compare ratings of grammatical and ungrammatical sentences statistically, we ran a series of mixed-effect linear

regressions. To ensure that there were no differences across the counterbalancing conditions, we used a model with main effects and interactions of grammaticality, counterbalancing condition, and block as fixed effects, with by-subject random intercepts and by-subject random slopes for grammaticality. We found that the three-way interaction was not significant,  $\chi^2(2) = 0.06$ ,  $p = .971$ . Counterbalancing group also did not interact individually with grammaticality or block (all  $ps > .05$ ), so we collapsed across counterbalancing groups. We then ran an analysis with grammaticality and block, and their interaction, as fixed effects, and by-subject random intercepts and by-subject random slope for grammaticality. The interaction was not significant,  $\chi^2(2) = 5.36$ ,  $p = .069$ , so we removed the interaction from the model and reran the analysis. With both grammaticality and block as the fixed effect, and subject as random intercept and by-subject random slope for grammaticality, we found that block was not significant,  $\chi^2(2) = 1.51$ ,  $p = .471$ , but grammaticality was significant suggesting that participants were able to learn the adjacent dependency ( $\beta = 0.398$ ,  $z = 3.79$ ,  $p < .001$ ). We conclude that participants were successful in learning the adjacent dependency in the *in-phase* condition (see Fig. 2).

For the *out-of-phase* condition, the mean rating of grammatical sentences was 2.810, and the mean rating of ungrammatical sentences was 2.762. To compare ratings of grammatical and ungrammatical sentences statistically, we ran mixed-effect linear regressions on the data. To ensure that there were no differences across the counterbalancing conditions, we fit a Grammaticality  $\times$  Counterbalancing  $\times$  Block interaction as the fixed effect, and subject as a random intercept and a by-subject random slope for grammaticality as a random effect. We found that the three-way interaction was not significant,  $\chi^2(2) = 0.39$ ,  $p = .821$ . Further analysis showed that counterbalancing group does not interact with grammaticality or block, either (all  $ps > .05$ ), and we collapsed the counterbalancing group for the subsequent analysis. We then ran an analysis with grammaticality and block, and their interaction, as fixed effects, and by-subject random intercepts and by-subject random slopes for grammaticality. The interaction was not significant,  $\chi^2(2) = 2.40$ ,  $p = .301$ , so we removed it from the model. With grammaticality and block as the fixed effect, with subject as random intercept, and by-subject random slope for grammaticality, neither block,  $\chi^2(2) = 3.70$ ,  $p = .157$ , nor grammaticality ( $\beta = 0.048$ ,  $z = 0.84$ ,  $p = .402$ ) was significant. Thus, we found no evidence that participants learned the adjacent dependencies in the *out-of-phase* condition (see Fig. 3).

Due to the theoretical importance of the null effect in the *out-of-phase* condition, we applied a Bayesian approach to the results of both experiments in order to compare how well the null hypothesis fit each and to demonstrate that the null effect of grammaticality was not due to lack of power. Following the procedure in Gallistel (2009), the analysis showed that the

<sup>3</sup> The data for this paper is available at <https://osf.io/x9msp/>.



**Fig. 2** Means for grammatical and ungrammatical test items, and difference scores for each subject in the *in-phase* condition, collapsed over counterbalancing conditions and blocks. **a** Each circle represents the mean rating of a subject for all grammatical and ungrammatical items, with a solid line indicating the mean for each item type. **b** Each

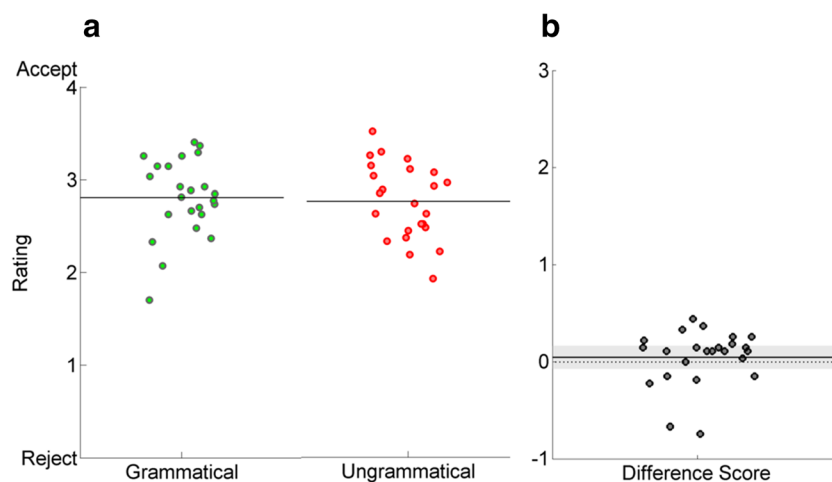
circle represents the difference between mean ratings (grammatical – ungrammatical) for each subject, with the solid line showing the mean difference, and shadows showing 95% confidence intervals around the mean. The dotted line at zero represents chance

odds for the null in the *out-of-phase* condition is 19.4278, which is considered strong evidence for the null hypothesis (Jeffreys, 1961). For comparison purposes, the odds for the null in the *in-phase* condition is 2.4034e-09, which is evidence for the alternative hypothesis.

To compare the *in-phase* and *out-of-phase* conditions, we ran a mixed-effect linear regression with the rating data as the dependent variable. Item type (grammatical vs. ungrammatical) and condition (*in-phase* and *out-of-phase*) and their interaction served as fixed effects, with a by-subject random intercept and slope for grammaticality. The interaction was significant,  $\chi^2(2) = 9.23$ ,  $p = .002$ , suggesting that the two conditions had different learning outcomes.

## General discussion

In this study, we deployed a method of inducing top-down expectations that serve as grouping information, without manipulating the prosody of the artificial language that carried the statistical information or modifying the bottom-up signal in any other way. We found that learners succeeded in learning the statistical relations between elements when they were presented in phase with the entrained rhythm, suggesting that participants can learn from statistical cues within sub-sequences. However, when the elements making up the adjacent dependencies were bracketed into different sub-sequences (via introducing a subtle change to the English sentences so



**Fig. 3** Means for grammatical and ungrammatical test items, and difference scores for each subject in the *out-of-phase* condition collapsed over counterbalancing conditions and blocks. **a** Each circle represents the mean ratings of a subject for grammatical and ungrammatical items, with a solid line indicating the mean for each

item type. **b** Each circle represents the difference between mean ratings (grammatical – ungrammatical) for each subject, with the solid line showing the mean difference, and shadows showing 95% confidence intervals around the mean. The dotted line at zero represents chance



that the dependent words, without manipulating the sequences of nonsense words), participants failed to learn the adjacent dependencies.

The fact that learning failed in the *out-of-phase* condition is particularly noteworthy, as the adjacent dependencies had a conditional probability of 1. Such strong dependencies have been shown in many studies to be very easy to learn (e.g., Saffran et al., 1996). Our claim is that the top-down expectations act as a filter that shapes the “intake” of the statistical learning mechanisms (Gagliardi & Lidz, 2014; Mintz, Wang, & Li, 2014). Under this view, the representation the learner builds is shaped by existing knowledge and expectations. In our study, the English sentences preceding the artificial ones were all four words long. This shaped the processing of the subsequent artificial words, such that subjects bracketed them into four-word-long sequences and only consider the relationships within the bracketed artificial words. Thus, learning the dependencies took place only when the elements making up the dependencies were within bracketed sub-sequences.

This is importantly different from past studies in which two sources of bottom-up information, such as prosodic information and statistical information, were presented. For example, in a word segmentation study, Shukla et al. (2007) trained subjects on sequences of syllables with overlaid prosodic contours—a form of bottom-up grouping information. They manipulated the alignment of prosodic units—intonational phrases (IPs)—with trisyllabic sequences of syllables with word-internal transitional probabilities (TPs) of 1. In a two-alternative forced-choice test, subjects selected these high-TP “words” over “nonwords”—sequences that had not occurred—only when the words had been contained within an IP, not when they had straddled an IP boundary. Nevertheless, the researchers further showed that subjects stored and could access the statistical information (high-TP words), even when the statistical grouping information conflicted with the prosodic information (Shukla et al., 2007, Experiment 4). They concluded that the two sources of bottom-up information are computed and stored in parallel, and that prosodic information suppresses the access of statistical information in the auditory domain. Those findings, considered with the findings we report here, suggest that the way bottom-up and top-down sources of bracketing information influence statistical processing differ.

Mattys and Bortfeld (2016) provided a framework in which local cues (such as stress and statistics) are initially predominant, but gradually downplayed as lexical knowledge and more global cues (such as distal prosodic cues and contextual information) play a more significant role as a developmental trajectory, which is consistent with the novel framework developed in this paper.

While adjacent dependencies have been demonstrated to be easy to learn in most studies, nonadjacent dependencies have been theorized to be difficult to learn (Gomez, 2002; Newport

& Aslin, 2004; cf. Wang, Zevin, & Mintz, 2019). We recently conducted a similar grammatical entrainment study with non-adjacent dependency learning (Wang et al., 2017) and found that under entrainment influences similar to the current experiment, learners succeed at nonadjacent dependency learning if the elements making up the nonadjacent dependencies are bracketed in the same sub-sequences and fail if the elements are bracketed in different sub-sequences. In other words, we found that bracketing has the same effect on adjacent and nonadjacent dependency learning, in that learning occurs only when the adjacency appears within the same sub-sequence. Together, these studies show that bracketing of the continuous sequence produces the sub-sequences that learners have access to, such that the bracketing procedure largely determines what can and cannot be learned, rather than the type of linguistic dependencies to be learned.

We now would like to address an alternative explanation of the learning failure in the *out-of-phase* condition. Recall that the phase was shifted by relocating the first word of the first sentence in a sequence of English sentences to the end of the last sentence (see Fig. 1). This resulted in an English word (*Kate*, in Fig. 1) starting a “sentence” where the rest of the words are artificial. Could this have simply disoriented or distracted subjects so that they were not paying attention to the subsequent material, including the dependencies? We believe that there is evidence against this alternative, in the study we discussed in the previous paragraph (Wang et al., 2017). In one experiment in that study, we produced out-of-phase stimuli as we did here, but in another experiment, we substituted a new nonsense word (i.e., replacing *Kate*) at the beginning of the artificial language material. This implemented the same phase shift, but without mixing English with artificial words in the bracketing cycle. There, we found the same disruption of dependency learning, indicating that the effect was not due to the mixing of English with artificial words. Given that finding, we have no reason to believe that the presence of an English name at the start of a sequence of artificial words would be any more disruptive here.

A final question of interest is whether there is, in addition to or separate from, a role of syntax/semantics of the English sentences on the learning of adjacent dependencies. Despite that this is an intriguing possibility, we believe that it is unlikely based on both theoretical and empirical grounds. Theoretically, the adjacent dependencies are positioned in the third and fourth positions of these parsed “sentences.” The English words at the third and fourth positions are the pronouns and prepositions (e.g., *this up*), and do not make up a linguistic dependency (as mentioned above, this structure was specifically chosen for this reason). If one is to look for correspondence between the third and fourth positions in the artificial language in both the *in-phase* and the *out-of-phase* conditions, they correspond to English phrases such as “this up” in the *in-phase* condition and “up John” in the *out-of-*

*phase* condition, neither of which are syntactic constituents, nor are the words in either sequence syntactically dependent. Moreover, previous work (Wang et al., 2017) tested nonadjacent dependency learning with repeating sentences that had a phrase structure that either highlighted the nonadjacent dependency (i.e., John puts these up), or a phrase structure that did not contain nonadjacent dependencies (i.e., John cuts through these). The results did not differ as a result of the different phrase structures in these experiments. The only factor that mattered was whether the sub-sequences from sentential level expectations contained the to-be-learned dependencies, much like the result of the current study.

In conclusion, our study showed that top-down grouping information influences statistical learning. Even adjacent dependencies with transitional probability of one are only learned when they are contained within, versus across, the sub-sequences. In other words, when grouping information is available, learning takes place between elements within sub-sequences rather than all the elements in the continuous sequence. This limits the amount of linguistic computations that need to be performed, providing a domain for statistical learning to operate over.

**Open practice statement** The reported experiment was not preregistered. The data reported in this paper, as well as sample stimuli, can be found at: <https://osf.io/x9msp/>

## References

- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321–324.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–347.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, 51(4), 523–547.
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016a). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1), 158–164.
- Ding, N., Patel, A., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2016b). Temporal modulations reveal distinct rhythmic properties of speech and music. *BioRxiv*, 059683.
- Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, 18(3), 254–260.
- Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, 90(1), 58–89.
- Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, 116(2), 439.
- Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436.
- Jeffreys, H. (1961). *The theory of probability*. New York: Oxford University Press.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4), 548–567.
- Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, 12(1), 131–141.
- Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, 14(6), 1323–1329.
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134(4), 477.
- Mattys, S. L., & Bortfeld, H. (2016). Speech segmentation. In M. G. Gaskell, & J. Mirkovic (Eds.) *Speech perception and spoken word recognition current issues in the psychology of language* (pp. 55–75). Abingdon, Oxon: Psychology Press.
- Mintz, T. H., Wang, F. H., & Li, J. (2014). Word categorization from distributional information: Frames confer more than the sum of their (bigram) parts. *Cognitive Psychology*, 75, 1–27.
- Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, 108(1), 271–280.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19, 498–550.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language*, 28(3), 360–374.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127–162.
- Palmer, S. D., Hutson, J., White, L., & Mattys, S. L. (2019). Lexical knowledge boosts statistically-driven speech segmentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 139.
- 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.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Shukla, M., Nespore, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54(1), 1–32.
- Wang, F. H., Zevin, J., & Mintz, T. (2017). Grammatical bracketing determines learning of non-adjacent dependencies. *Journal of Experimental Psychology: General*.
- Wang, F. H., Zevin, J., & Mintz, T. H. (2019). Successfully learning non-adjacent dependencies in a continuous artificial language stream. *Cognitive Psychology*, 113, 101223.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.