Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration

Esther van den Bos [a],[*],[1], Morten H. Christiansen [b],[*], Jennifer B. Misyak [b]

[a] Department of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands
[b] Department of Psychology, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

ABSTRACT

Previous studies have indicated that dependencies between nonadjacent elements can be acquired by statistical learning when each element predicts only one other element (deterministic dependencies). The present study investigates statistical learning of probabilistic nonadjacent dependencies, in which each element predicts several other elements with a certain probability, as is more common in natural language. Three artificial language learning experiments compared statistical learning of deterministic and probabilistic nonadjacent dependencies. In Experiment 1, participants listened to sequences of three non-words containing either deterministic or probabilistic dependencies between the first and the last non-words. Participants exposed to deterministic dependencies subsequently distinguished correct sequences from sequences that violated the nonadjacent dependencies; participants exposed to probabilistic dependencies did not. However, when visual (Experiment 2) and phonological cues (Experiment 3) were added participants learned both kinds of dependencies, demonstrating statistical learning of probabilistic nonadjacent dependencies when additional cues are present.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Remote (or nonadjacent) dependencies are abundant in natural language. In English, for example, linguistic material may intervene between auxiliaries and inflectional morphemes (e.g., *is cooking*, *has traveled*) or between subject nouns and verbs in number agreement (*the books on the shelf are dusty*). More complex relationships to surface forms may be found in nonadjacent dependencies in anaphoric reference (e.g., *John went to the store where he bought some milk*) as well as between *wh*-questions and the semantic expectations for the answer (e.g., *Who/What did you see?* where the answer is expected to be a person for *who* and an object or event for *what*).

Children's sensitivity to nonadjacent dependencies in language emerges gradually. The general pattern of acquisition indicates that nonadjacent dependencies apparent in the surface structure of sentences are acquired earlier than more abstract nonadjacent dependencies (such as that involved in anaphoric reference). Santelmann and Jusczyk (1998) showed that 18-month-olds are sensitive to violations of nonadjacent dependencies between *is* and *–ing* in comprehension, and the use of the present progressive morpheme *–ing* also shows up early in production (though initially without the appropriate dependency relation to the auxiliary *is*; Brown, 1973). On the other hand, even though *wh*-questions show up in children's productions as early as 17 months, they tend to follow formulaic patterns (see O'Grady, 1997, for discussion). Even after children have otherwise mastered subject-noun/verb agreement around 2–2.5 years of age, they still produce incorrect *wh*-questions with agreement violations (such as, *What colour is these?*; Radford, 1990). Moreover, they also have problems responding correctly to *wh*-questions

* Corresponding authors. Fax: +31 71 5273619 (E. van den Bos).
  E-mail addresses: bosejvanden@fsw.leidenuniv.nl (E. van den Bos), christiansen@cornell.edu (M.H. Christiansen).
[1] Esther van den Bos was supported by a grant from the Niels Stensen Stichting.

involving a direct object *wh*-word and a non-copular verb (such as, *What did mummy say?* to which a 21-month-old responded *Mummy*; Radford, 1990).

Although children eventually acquire these complex relationships, it is not yet clear exactly how children (and adults) learn nonadjacent dependencies. One suggestion is that statistical learning—the discovery of structure by way of statistical properties of the input—may play an essential role in learning dependencies between different elements in language (for reviews, see e.g., Gómez & Gerken, 2000; Misyak, Goldstein, & Christiansen, in press; Redington & Chater, 1998; Romberg & Saffran, 2010; Saffran, 2003). For example, it has been shown that participants become sensitive to transitional probabilities between syllables in a speech stream (Saffran, Aslin, & Newport, 1996) and between fixed groups of tones (Saffran, Johnson, Aslin, & Newport, 1999) and individual tones in a musical sequence (Saffran, 2002). In the visual domain, participants become sensitive to transitional probabilities between shapes in a visual sequence (Saffran, 2002) as well as to probabilities of co-occurrence of objects in a scene (Fiser & Aslin, 2005). Although the structure consisted of dependencies between adjacent elements in these studies, several other studies have shown that dependencies between nonadjacent elements can be acquired by statistical learning as well (e.g., Endress & Bonatti, 2007; Gebhart, Newport, & Aslin, 2009; Gómez, 2002; Newport & Aslin, 2004; Onnis, Monaghan, Richmond, & Chater, 2005; Pacton & Perruchet, 2008; Peña, Bonatti, Nespor, & Mehler, 2002).

Some studies have demonstrated that participants can learn nonadjacent dependencies between (artificial) classes of syllables in a speech stream (e.g., Endress & Bonatti, 2007; Onnis et al., 2005; Peña et al., 2002). Other studies have shown that participants can learn nonadjacent dependencies between individual elements (e.g., Gómez, 2002; Newport & Aslin, 2004; Pacton & Perruchet, 2008). The focus has been on statistical learning of nonadjacent dependencies involving surface level cues rather than on more abstract dependencies (e.g., comparable to anaphoric reference). The present study will also be concerned with this more tractable problem of learning nonadjacent dependencies from surface level cues. Although such learning may not be considered crucial from a generativist perspective, it provides valuable information on the degree to which syntactical structures can be learned from experience. We expect that a thorough understanding of learning nonadjacent dependencies from surface level cues can provide an important scaffolding for future work tackling the learning of more abstract nonadjacent structure (regardless of whether the two processes prove to be similar or different).

There is some evidence that statistical learning may be less reliable for nonadjacent dependencies than for adjacent dependencies (though see Endress & Bonatti, 2007). Participants in a speech segmentation study (Newport & Aslin, 2004) learned to discriminate between words in an artificial language and speech fragments spanning word boundaries from listening to a continuous stream of synthetic speech. However, this finding was only obtained when the words were characterized by dependencies between consonants, spanning the vowels, or by dependencies between vowels, spanning the consonants. When the words were characterized by nonadjacent dependencies between syllables, participants failed to discover the words. Moreover, another speech segmentation study (Bonatti, Peña, Nespor, & Mehler, 2005) showed that nonadjacent dependencies between vowels were only acquired when a word containing a particular nonadjacent dependency between the vowels was regularly followed by another word containing the same nonadjacent dependency.

Newport and Aslin (2004) proposed that the dependent elements have to be perceptually similar to enable statistical learning of nonadjacent elements in a continuous stream. In line with this proposal, statistical learning of nonadjacent dependencies between syllables was demonstrated when the dependent syllables were phonologically similar, i.e., when the dependent syllables started with a plosive and the intervening syllable started with a continuant or vice versa (Onnis et al., 2005). Similarly, statistical learning of nonadjacent dependencies in musically meaningless sequences of tones only occurred when nonadjacent tones were similar in pitch range or timbre, while adjacent tones were different (Creel, Newport, & Aslin, 2004). However, similarity was not required to learn nonadjacent dependencies in musically meaningful sequences, in which the difference in pitch between two nonadjacent tones was bridged by an intervening tone (Endress, 2010).

Gómez (2002) suggested that statistical learning is tuned to adjacent dependencies by default, but that nonadjacent dependencies are learned when the predictive relationships between nonadjacent elements are stronger than those between adjacent elements. She observed that this is the case in natural language, where nonadjacent dependencies tend to involve elements from small classes, like functional morphemes, separated by elements from large classes, like nouns and verbs (e.g., <u>is</u> work<u>ing</u>, <u>is</u> play<u>ing</u>, etc.). An artificial language learning experiment provided evidence for the influence of the relative strength of the predictive relationships between adjacent and nonadjacent elements. Adults and 18-month-old children, who listened to sequences of three non-words containing nonadjacent dependencies between the first and the third, were better at recognizing those dependencies when they had been exposed to 24 elements in the middle than when they had been exposed to a small set of intervening elements. A subsequent study demonstrated that high variability of the intervening element also enabled participants to segment a continuous speech stream into non-words characterized by nonadjacent dependencies between syllables (Onnis, Monaghan, Christiansen, & Chater, 2004). Furthermore, statistical learning of nonadjacent dependencies was also enhanced when adjacent dependencies were uninformative because the material separating the dependent elements was invariant instead of highly variable (Onnis, Christiansen, Chater, & Gómez, 2003).

Pacton and Perruchet (2008) manipulated task demands to guide attention to either adjacent or nonadjacent dependencies. In their experiments, participants were instructed to detect a target in a sequence of digits and to subtract either the two digits following each target or the two digits surrounding it. All sequences contained an

adjacent dependency and a nonadjacent dependency. On each trial, a target focused attention on the adjacent dependency for participants subtracting the subsequent digits and on the nonadjacent dependency for participants subtracting the surrounding digits. Participants who subtracted the digits following a target only learned the adjacent dependency, whereas participants who subtracted the digits surrounding a target only learned the nonadjacent dependency.

Thus, previous research suggests that dependencies between nonadjacent elements can be learned reliably from exemplars when task demands or cues such as similarity and the relative (in)stability of the statistical context constrain the learner's hypothesis space[2] to nonadjacent structures. However, the evidence is currently limited to deterministic nonadjacent dependencies, in which each element is always followed eventually by only one other element. Probabilistic nonadjacent dependencies, in which several elements have a certain probability to eventually follow each element, may be more difficult to learn. Firstly, the predictive relationship in a probabilistic nonadjacent dependency is, by definition, weaker than in a deterministic one. Secondly, implicit learning studies have provided evidence that higher order dependencies, in which the next element of a sequence can be predicted on the basis of multiple previous elements, are difficult to learn.

In one such study (Cohen, Ivry, & Keele, 1990), participants performed a Serial Reaction Time (SRT) task. On each trial, they were presented with a stimulus appearing in one of six locations on the computer screen. Their task was to press the key on the keyboard that corresponded to the location on the screen where the stimulus had appeared. For some participants, a stimulus in one location (e.g., location 2) could only be followed by a stimulus in one other location (e.g., location 3). For other participants, a stimulus in one location (e.g., location 2) could be followed by a stimulus in two other locations (e.g., locations 3 and 4), depending on the location by which it was preceded (e.g., 523 and 124). After several training blocks, there were two test blocks in which some stimuli appeared in random locations. The first group showed evidence of sequence learning by reacting faster to stimuli that appeared in predictable as opposed to random locations, even when participants had to perform an additional tone counting task. The second group, in contrast, did not show evidence of sequence learning when they had to perform the additional tone counting task. As the location of the next stimulus was uniquely determined by the two previous locations, this indicates that participants did not learn higher order dependencies under dual task conditions. In addition, implicit learning of higher order dependencies in probabilistic structures has been shown to require extended exposure (Cleeremans & McClelland, 1991) and it may fail when the structure is highly complex (Van den Bos & Poletiek, 2008).

The question of whether or not probabilistic nonadjacent dependencies can be acquired by statistical learning is particularly important, because the nonadjacent dependencies that occur in natural language tend to be probabilistic. For example, in languages such as Dutch, English, and German, the perfect participle is formed differently for weak verbs and strong verbs. In German, the perfect participle of a weak verb is formed by the prefix *ge-*, the word stem and the suffix *-t* (e.g., *Ich habe gesagt* ('I have said')), but for strong verbs the perfect participle is formed by the prefix *ge-*, the word stem (with vowel change) and the suffix *-en* (*Ich habe gesprochen* ('I have spoken')).

Whereas previous work on statistical learning of nonadjacent dependencies has focused on deterministic dependency relations, in which the dependent elements invariably occur together, the present study investigates whether this type of learning can be extended to probabilistic nonadjacent dependencies, in which multiple elements can occur given a previous element. In Experiment 1, participants were presented with nonadjacent dependencies instantiated in sequences of non-words. In two subsequent experiments, the sequences of non-words were complemented with visual and phonological cues.

Experiment 1

In this experiment, participants listened to sequences consisting of three non-words, which contained either a probabilistic or a deterministic nonadjacent dependency between the first and the third non-word. For the deterministic dependencies, each of the four initial non-words was paired with a single unique final non-word. For the probabilistic dependencies, two initial non-words were each paired with two final non-words. This situation parallels subject–verb agreement in the present and the past tense for regular verbs in Dutch, where each noun may be followed by a verb with two possible endings (e.g., *de kikker kwaakt* ('the frog croaks'), *de kikker kwaakte* ('the frog croaked'), *de kikkers kwaken* ('the frogs croak') and *de kikkers kwaakten* ('the frogs croaked')). To replicate the high variability condition of Gómez (2002), in which participants were most successful at learning deterministic nonadjacent dependencies, 24 bisyllabic non-words were used in the middle, while the initial and final non-words were monosyllabic.

Method

Participants

Forty-eight undergraduate students from Cornell University participated in the experiment. They received either course credit or $5 for their participation. The data from four participants could not be included in the analyses, because of computer failure (2), experimenter error (1) or external disturbance (1) during the experiment. Of the remaining participants (16 male, 28 female, 18–23 years of age), 22 were assigned to each condition.

Materials

Participants were exposed to sequences of three spoken non-words. The last element of each sequence depended on the first, while the second was variable. There were four

---
[2] The term 'hypothesis space' is not meant to imply that the learner consciously holds or tests hypotheses.

possible combinations of initial and final elements for each type of nonadjacent dependency. The deterministic nonadjacent dependencies were instantiated in four initial and four final non-words. Each initial non-word occurred with one final non-word (a_1Xb_1, a_2Xb_2, a_3Xb_3, a_4Xb_4), so that the conditional probability of the final element given the initial element was 1. The probabilistic nonadjacent dependencies were instantiated in two initial and four final non-words. Each initial non-word occurred with two final non-words (a_1Xb_1, a_1Xb_2, a_2Xb_3, a_2Xb_4), so that the conditional probability of the final element given the initial element was .50. Twenty-four non-words were used as middle elements and presented with each of the four dependencies, leading to 96 unique sequences in each condition.

In the test phase, participants were presented with 16 sequences. Eight test sequences contained the dependencies that a participant had been trained on, while the other eight violated the dependencies. The violations of the dependencies were created by pairing the initial non-words with a final non-word with which it had not been presented in the training phase (a_1Xb_2, a_2Xb_3, a_3Xb_4, a_4Xb_1 for the deterministic nonadjacent dependencies and a_1Xb_3, a_1Xb_4, a_2Xb_1, a_2Xb_2 for the probabilistic nonadjacent dependencies). Thus, knowledge of which non-words occurred in a certain position (initially and finally) in the sequence was insufficient to detect the violation of a dependency. Each of the eight combinations of initial and final elements was presented with two different middle elements. Apart from that restriction, the middle element was randomly selected from the set of 24 non-words on each trial.

The set of middle elements consisted of the non-words: *balip, benez, cloval, coomo, deecha, feenam, fengle, gensim, gople, hiftam, loga, laeljeen, malsig, nilbo, plizet, puser, roosa, skigger, suleb, taspu, tisso, vamey, wadim, yafta*. The non-words used as initial and final elements were: *dak, meep, pel, rud, sig, tood, vot, zoet*. They were randomly assigned to a_{1-4} and b_{1-4} for each participant, to control for possible effects of phonological properties of the non-words on learning the dependencies. The non-words assigned to a_3 and a_4 were not used for the probabilistic dependencies. The non-words were recorded by a female native speaker of American English. The recordings were edited so that the duration was 500 ms for the monosyllabic non-words and 600 ms for the disyllabic non-words. (This allowed us to present visual cues exactly simultaneously with the non-words in Experiment 2) The non-words were combined into sequences with 250 ms pauses between the elements. There was a 750 ms pause between successive sequences.

Procedure

In the training phase, participants were exposed to four blocks of 96 sequences of non-words, presented in random order. They were instructed to listen attentively, because their knowledge of the sequences would later be tested. The training phase took about 19 min. At the beginning of the test phase, participants were informed that the sequences of non-words had been formed according to certain rules. It was announced that they would now hear eight sequences that followed the same rules and eight

sequences that violated the rules. Participants had to judge for each sequence whether or not it followed the rules. They were instructed to press the green button on the keyboard if they thought that it followed the rules and the red button if they thought that it violated the rules. In all, the experiment took about 25 min.

Results and discussion

An independent samples *t*-test (not assuming equal variances) showed that the proportion of correct grammaticality judgments was higher for the deterministic dependencies than for the probabilistic dependencies ($t(39.3) = 2.161$, $p = .037$). This suggests that people are better at statistical learning of nonadjacent dependencies when one element always follows a certain other element after a highly variable intervening element than when two possible elements can follow a certain element after a highly variable intervening element. This result for probabilistic nonadjacent dependencies is in line with previous findings that higher-order dependencies in a probabilistic structure are relatively difficult to learn (e.g., Cleeremans & McClelland, 1991; Cohen et al., 1990; Van den Bos & Poletiek, 2008).

The difference in performance on deterministic and probabilistic nonadjacent dependencies was not due to a difference in response bias. The mean endorsement rate did not differ between deterministic ($M = 9.5$, $SD = 2.2$) and probabilistic ($M = 10.2$, $SD = 3.1$) nonadjacent dependencies ($t(42) < 1$, 95% CI difference = -2.4–0.9). Also, there was no significant difference between deterministic ($n = 0$) and probabilistic ($n = 2$) nonadjacent dependencies ($\chi^2(1) = 2.095$, $p = .148$) in the number of participants who endorsed all items, a response pattern that may occur when participants only learn which words occur in each position in the sequence.

The mean proportion of correct grammaticality judgments for the deterministic nonadjacent dependencies was significantly above chance ($M = .696$, $SD = .245$, $t(21) = 3.753$, $p = .001$), confirming that participants had acquired knowledge about the dependencies. Performance was not as high as in the original study by Gómez (2002), but it was comparable to the levels of performance observed in subsequent replications (e.g., Misyak & Christiansen, 2012; Onnis et al., 2003), even though participants in the present study were presented with four instead of three dependencies and the unique sequences were repeated less often.

For the probabilistic dependencies, however, the mean proportion of correct grammaticality judgments failed to reach significance ($M = .554$, $SD = .187$, $t(21) = 1.352$, $p = .191$, 95% CI difference = $-.029$–.137). This finding is remarkable, because previous studies using this paradigm with deterministic nonadjacent dependencies tended to find some evidence of learning, even under difficult conditions. For example, although participants who had been exposed to a small set of intervening elements in the study by Gómez (2002) did less well on the grammaticality judgment test than participants who had been exposed to 24 intervening non-words, their performance was still above chance in some conditions. Similarly, participants have

been shown to learn deterministic nonadjacent dependencies spanning three elements when each of the intervening elements was drawn from a small set of non-words and the adjacent dependencies were, as a result, relatively strong (Misyak & Christiansen, 2007).

The difference between the two conditions was therefore further investigated by comparing the number of participants who learned at least one nonadjacent dependency. As noted by Hsu, Tomblin, and Christiansen (in preparation), participants may have learned some of the nonadjacent dependencies even when their overall performance is at chance. A dependency was considered to be learned when a participant responded correctly to both the grammatical item and the nongrammatical item with the same initial non-word in both test blocks. In the probabilistic condition, the number of correct dependencies was averaged over the two possible mappings of grammatical and ungrammatical items with each initial non-word. Fifteen participants learned at least one dependency in the deterministic condition. In the probabilistic condition, only 3 participants learned at least one dependency. The difference was significant ($\chi^2(1) = 13.538$, $p < .001$). Although more people might learn probabilistic nonadjacent dependencies when the variability of the intervening material is increased further, we chose to explore whether statistical learning of probabilistic nonadjacent dependencies could be enhanced when additional cues are presented.

Experiment 2

A well-tried method to enhance learning of complex aspects of an artificial language is to supply a visual reference field (e.g., Meier & Bower, 1986; Moeser & Bregman, 1972, 1973; Mori & Moeser, 1983; Nagata, 1977; Valian & Coulson, 1988). In Experiment 2, we presented the sequences of non-words in the context of moving colored shapes. The initial non-words were paired with colors, the middle non-words were paired with shapes and the final non-words were paired with continuous and single movements (with the Dutch example of subject–verb agreement in the present and the past tense in mind). Thus, the auditory sequences contained a nonadjacent dependency between the initial and the final non-word and the visual illustrations contained a dependency between the color and the movement of the shape. As a consequence, the initial and the final non-word of a sequence were both paired with the same color and movement, which may highlight the dependency between them. We hypothesized that the consistent association of non-words with visual referents would enhance statistical learning of probabilistic nonadjacent dependencies. We tested this hypothesis by providing participants with visual cues that were either congruent or incongruent with the auditory nonadjacent dependencies.

Method

Participants

The participants in this experiment were 92 undergraduate students from Cornell University who had not participated in Experiment 1. They received either course credit or $5 for their participation. The data from 4 participants were discarded, because of computer failure (1), experimenter error (2) or color blindness (1). Of the remaining 88 participants (27 male, 61 female, 18–35 years of age), 22 were assigned to each of four conditions (congruent vs. incongruent visual cues × deterministic vs. probabilistic nonadjacent dependencies).

Materials

The auditory stimuli were the same as in Experiment 1. The visual stimuli consisted of moving colored shapes. There were 96 unique bitmap images, produced by creating blue, green, red and yellow versions of 24 abstract shapes from Emberson, Conway, and Christiansen (2011) and Fiser and Aslin (2002). The four possible movements were: blinking once, blinking twice, moving back and forth once, moving back and forth twice.

In the conditions with congruent illustrations, the visual stimuli were systematically related to the auditory stimuli (see Table 1, for examples of the relationship between auditory and visual stimuli for both types of dependencies). In the condition with deterministic dependencies, each color was randomly assigned to an initial non-word, each shape was randomly assigned to a middle non-word and each movement was randomly assigned to a final non-word. In the condition with probabilistic dependencies, two colors and one type of movement (blink vs. back and forth) were randomly selected for each participant. The colors were randomly assigned to the initial non-words a_1 and a_2. 'Moving once' and 'moving twice' were randomly assigned to the final non-words associated with a_1 (i.e., b_1 and b_2), and the final non-words associated with a_2 (i.e., b_3 and b_4). The shapes were randomly assigned to the non-words in the middle.

In the conditions with incongruent visual illustrations, colors, shapes and movements were determined randomly on each trial, so that there was no systematic relationship between auditory and visual stimuli. Only two colors were used in the condition with probabilistic nonadjacent dependencies (randomly selected for each participant).

Procedure

The procedure for the training phase was similar to Experiment 1, but the sequences of non-words were accompanied by moving colored shapes. At the beginning of the experiment, the participants were informed that they would be presented with auditory sequences of three non-words and visual illustrations. They were instructed to listen attentively, because their knowledge of the sequences would later be tested in the absence of visual illustrations. The procedure for the test phase was identical to Experiment 1 (with no visual illustrations presented during testing).

Results and discussion

The results are shown in Fig. 1. A univariate analysis of variance (ANOVA) with type of nonadjacent dependency (deterministic vs. probabilistic) and type of illustration (congruent vs. incongruent) as independent variables and

Table 1
Example of the relationship between auditory and visual stimuli in the congruent condition for deterministic and probabilistic nonadjacent dependencies.

Deterministic				Probabilistic			
Sequence	Color	Shape	Movement	Sequence	Color	Shape	Movement
dak balip meep	blue	1	blink 1x	dak balip meep	blue	1	blink 1x
dak . . . meep	blue	. . .	blink 1x	dak . . . meep	blue	. . .	blink 1x
dak yafta meep	blue	24	blink 1x	dak yafta meep	blue	24	blink 1x
pel balip rud	green	1	blink 2x	dak balip rud	blue	1	blink 2x
pel . . . rud	green	. . .	blink 2x	dak . . . rud	blue	. . .	blink 2x
pel yafta rud	green	24	blink 2x	dak yafta rud	blue	24	blink 2x
sig balip tood	red	1	back-forth 1x	pel balip tood	green	1	blink 1x
sig . . . tood	red	. . .	back-forth 1x	pel . . . tood	green	. . .	blink 1x
sig yafta tood	red	24	back-forth 1x	pel yafta tood	green	24	blink 1x
vot balip zoet	yellow	1	back-forth 2x	pel balip zoet	green	1	blink 2x
vot . . . zoet	yellow	. . .	back-forth 2x	pel . . . zoet	green	. . .	blink 2x
vot yafta zoet	yellow	24	back-forth 2x	pel yafta zoet	green	24	blink 2x
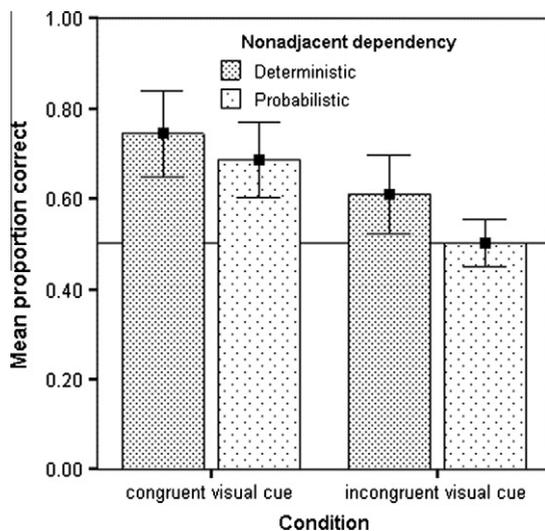
Fig. 1. Mean proportion of correct grammaticality judgments with 95% confidence intervals for deterministic and probabilistic nonadjacent dependencies presented together with congruent and incongruent visual cues.

the proportion of correct grammaticality judgments as the dependent variable showed significant main effects of type of nonadjacent dependency ($F(1,84) = 4.523$, $p = .036$) and type of illustration ($F(1,84) = 17.149$, $p < .001$). The interaction was not significant ($F(1,84) < 1$). The proportion of correct grammaticality judgments was higher for the deterministic dependencies ($M = .677$, $SD = .214$) than for the probabilistic dependencies ($M = .594$, $SD = .182$). Moreover, the proportion of correct grammaticality judgments was higher when the sequences of non-words were accompanied by congruent illustrations in the induction phase ($M = .716$, $SD = .202$) than when the sequences were accompanied by incongruent illustrations ($M = .555$, $SD = .168$).

When participants were presented with incongruent illustrations, performance was above chance for deterministic nonadjacent dependencies ($M = .609$, $SD = .195$, $t(21) = 2.628$, $p = .016$), but not for probabilistic nonadjacent dependencies ($M = .501$, $SD = 117$, $t(21) < 1$, 95% CI of

difference: $-.0514–.0523$). The pattern of results in this condition was similar to Experiment 1. The learning of deterministic nonadjacent dependencies in the auditory sequences may have been somewhat depressed by the concurrent presence of random visual stimuli, but it was not completely disrupted. Importantly, the results indicate that performance was not enhanced by the mere presence of visual illustrations, which, in principle, could have made the experiment more interesting to participants.

When participants were presented with congruent visual illustrations in the training phase, they scored above chance on the subsequent grammaticality judgment task for both deterministic ($M = .744$, $SD = .215$, $t(21) = 5.344$, $p < .001$) and probabilistic nonadjacent dependencies ($M = .688$, $SD = .189$, $t(21) = 4.654$, $p < .001$). The improvement in mean performance relative to Experiment 1 was similar for deterministic and probabilistic nonadjacent dependencies. An ANOVA on the proportion of correct grammaticality judgments did not show a significant interaction ($F(1,84) < 1$) between type of nonadjacent dependency (deterministic vs. probabilistic) and condition (no cue vs. congruent visual cue). However, the number of participants who learned at least one dependency only increased for probabilistic nonadjacent dependencies. An overall χ^2 test indicated that the number of participants who learned at least one dependency was not equally distributed over conditions and types of dependencies ($\chi^2(3) = 18.769$, $p < .001$). For probabilistic nonadjacent dependencies, the number of participants who learned at least one dependency was larger in the congruent visual cue condition ($n = 14$) than in the no cue condition ($n = 3$, $\chi^2(1) = 11.599$, $p = .001$). For deterministic nonadjacent dependencies, there was no difference between the no cue condition ($n = 15$) and the congruent visual cue condition ($n = 15$, $\chi^2(1) < 1$). The difference between deterministic and probabilistic nonadjacent dependencies was significant in the no cue condition ($\chi^2(1) = 13.538$, $p < .001$), but not in the congruent visual cue condition ($\chi^2(1) < 1$). Thus the results show that probabilistic nonadjacent dependencies can be learned when additional cues are provided.

Furthermore, the present experiment provides some insight into the role of the additional cues. As the visual cues were unavailable during the test phase, they must have

operated in the training phase. The participants' sensitivity to probabilistic nonadjacent dependencies seems to be due to the consistent associations between the dependent non-words and a visual referent. For example, many participants presented with probabilistic nonadjacent dependencies reported noticing that a shape of a particular color was paired with one initial non-word and two final non-words. Thus, as an alternative to making the dependent elements themselves more similar (Newport & Aslin, 2004), associating the dependent non-words with a single referent seems to further constrain the learner's hypothesis space and enhance statistical learning of nonadjacent dependencies.

The present experiment showed that probabilistic non-adjacent dependencies in an artificial language can be learned when they are presented together with visual cues. There is some evidence that visual information is used in processing and learning natural languages. For example, when participants were presented with sentences like "put the apple on the towel in the box," their eye movements indicated that they interpreted the phrase "on the towel" as a modifier of the noun "apple" if there were two apples present in the scene, whereas they often interpreted it as the goal location if there was only one apple (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). In addition, Yu (2006) has developed a computational model of language acquisition in which distributional information on the co-occurrence of words is integrated with distributional information about the co-occurrence of words and objects. Nevertheless, many aspects of natural languages are not consistently associated with visual information (Valian & Coulson, 1988). For example, Gleitman and Gleitman (1992) demonstrated that syntactic cues are far more useful than visual cues in learning the meanings of verbs. In Experiment 3, we investigated whether or not probabilistic nonadjacent dependencies could be learned in the presence of additional (nonvisual) cues within the artificial language.

Experiment 3A

In natural language, statistical regularities at the syntactic level are often associated with phonological regularities (see Monaghan & Christiansen, 2008, for a review). Although no single phonological property is perfectly predictive of word class, it has been demonstrated for English (Monaghan, Chater, & Christiansen, 2005; Monaghan, Christiansen, & Chater, 2007), Dutch, French and Japanese (Monaghan et al., 2007) that different phonological properties are associated with closed class and open class words and, within the latter, with nouns and verbs (see also Farmer, Christiansen, & Monaghan, 2006; Monaghan, Christiansen, Farmer, & Fitneva, 2010; Onnis & Christiansen, 2008). Several studies have indicated that such properties (in combination) can facilitate (artificial) language learning. Some of the phonological cues associated with nouns and verbs have been shown to enhance learning of two classes of non-words. The phonological cues were particularly useful for non-words that had been presented infrequently in the training phase, resulting in less reliable distributional cues

(Monaghan et al., 2005). In addition, 24-month-old children treated an unfamiliar non-word as a noun when it was accented, whereas they treated it as an adjective when it was de-accented (Thorpe & Fernald, 2006). Moreover, 2nd-graders used phonological cues when guessing whether a novel word refers to a picture of an object or a picture of an action (Fitneva, Christiansen, & Monaghan, 2009).

Of more direct relevance to the current experiment, an artificial language learning study with adults and 9–10-year-old children demonstrated a facilitative effect of phonological cues on learning subclasses of nouns corresponding to different genders. When a subset of the nouns in each class had a common ending, participants produced more correct sentences for new nouns than when the endings were counterbalanced over the subclasses (Brooks, Braine, Catalano, & Brody, 1993). Similarly, Spanish children who had to produce adjectives in an experiment with Spanish sounding non-words were shown to use distributional information provided by the Spanish determiners as well as phonological cues in the word endings to determine the gender of the adjective (Pérez-Pereira, 1991).

In Spanish, many masculine nouns and adjectives end in -o, while feminine determiners, nouns and adjectives often end in -a. However, the dependencies between word endings are probabilistic, because the phonological gender marking does not apply to all nouns and adjectives (e.g., los hombres honestos (the honest men), la casa grande (the big house); Pérez-Pereira, 1991). In Experiment 3, we investigated whether or not the addition of phonological cues would enhance statistical learning of probabilistic nonadjacent dependencies by instantiating the dependencies within non-words that shared a common ending.

Method

Participants

Ninety members of the Cornell University community participated in this experiment. None of them had participated in the previous experiments. Participants received either course credit or money as compensation for their time. Experimental participants were paid $5; control participants were paid $3. Data from two participants could not be included in the analysis, because of computer failure. Of the remaining 88 participants (33 male, 55 female, 18–52 years of age), 22 were assigned to each of the two experimental (deterministic vs. probabilistic nonadjacent dependencies) and two corresponding no-training control conditions.

Materials

The structure of the sequences of non-words was the same as in the previous experiments (deterministic: a_1Xb_1, a_2Xb_2, a_3Xb_3, a_4Xb_4; probabilistic: a_1Xb_1, a_1Xb_2, a_2Xb_3, a_2Xb_4). However, to instantiate a phonological cue involving common endings, bisyllabic non-words were used as initial and final elements in this experiment and monosyllabic non-words were used as middle elements. The use of monosyllabic non-words as middle elements kept the experiment similar to the previous experiments in two ways. First, this maintained the difference in the number of syllables between the initial and final

non-words on the one hand and the middle non-words on the other from Experiments 1 and 2, which may be a potential cue to the nonadjacent structure. Second, the use of monosyllabic middle elements resulted in a constant SOA between the initial and final elements across all experiments. The results of Emberson et al. (2011) suggest that increasing the temporal separation between initial and final elements any further may potentially hamper auditory statistical learning.

A new set of non-words was recorded from a female native speaker of American English. There were four subsets of initial and final elements, each containing three non-words with phonologically similar endings: *coomo, nilbo, tisso; gensim, pravin, wadim; loga, roosa, yafta* and *meeper, puser, skigger*. For the deterministic dependencies, one non-word from each subset was randomly selected as an initial element (e.g., a_1) and another non-word from that subset was randomly selected as the corresponding final element (e.g., b_1). For the probabilistic dependencies, two subsets were randomly selected. One non-word from each of these subsets was randomly selected as an initial element (e.g., a_1). The other non-words from these subsets were used as final elements (e.g., b_1 and b_2). The set of middle elements consisted of the non-words: *beel, bix, cav, dak, fis, foon, hes, jeen, jux, klor, koop, leeg, lum, neb, noob, pel, rauk, rud, sig, tam, tood, vot, zoet, zog*.

As before, the duration of the monosyllabic non-words was 500 ms and the duration of the disyllabic non-words was 600 ms. Pauses of 250 ms separated the non-words within a sequence and there was a 750 ms pause between successive sequences. As the sequences in Experiment 3 contained two disyllabic non-words, the total trial duration increased to 2950 ms.

Procedure

For the experimental groups, the procedure was the same as in Experiment 1. The control groups only participated in the test phase. Control participants were informed that they would be presented with 16 sequences of three non-words, 8 of which were formed according to certain rules and 8 of which violated those rules. Similar to the experimental participants, they were instructed to press the green button if they thought that a sequence followed the rules and the red button if they thought that it violated the rules. The experiment took about 25 min for the experimental groups and about 5 min for the control groups.

Results and discussion

The results are shown in Fig. 2 (Conditions: phonological cue, no training). A univariate ANOVA with type of nonadjacent dependency (deterministic vs. probabilistic) and condition (experimental vs. control) as the independent variables and the proportion of correct grammaticality judgments as the dependent variable did not show a significant effect of type of nonadjacent dependency ($F(1,84)$ < 1) or an interaction between type of nonadjacent dependency and condition ($F(1,84)$ < 1). However, there was a significant main effect of condition ($F(1,84) = 20.792$, $p < .001$). As expected, participants who had previously listened to sequences of non-words for 20 min ($M = .658$,
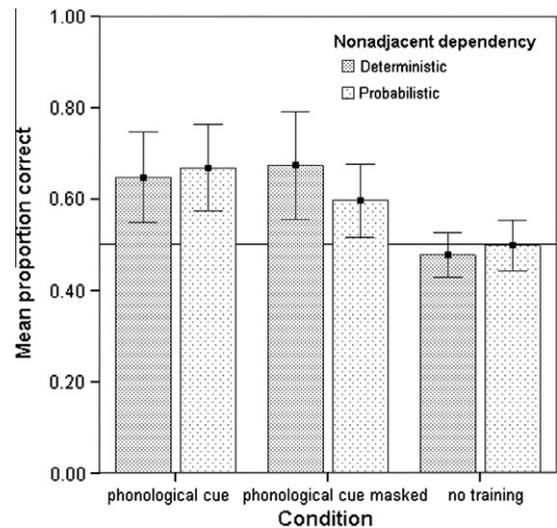

Fig. 2. Mean proportion of correct grammaticality judgments with 95% confidence intervals for deterministic and probabilistic nonadjacent dependencies when the phonological cue was present at test (3A), when the phonological cue was masked at test (3B: standard items) and in a no training control condition (3A).

$SD = .216$) did better on the grammaticality judgment task than participants who had not been exposed to the sequences of non-words prior to the test ($M = .488$, $SD = .116$).

One-sample *t*-tests showed that participants in the no-training control groups scored at chance for both deterministic ($M = .477$, $SD = .109$, $t(21) < 1$, 95% *CI difference*: $-.071–.026$) and probabilistic nonadjacent dependencies ($M = .498$, $SD = .123$, $t(21) < 1$, 95% *CI difference*: $-.056–.053$). This indicates that the phonological cue did not by itself lead participants to accept sequences in which the first and the third non-words rhymed and to reject sequences without rhyming as ungrammatical. Participants who were exposed to sequences of non-words in the training phase did better than control participants on deterministic ($M = .648$, $SD = .223$, independent samples *t*-test not assuming equal variances: $t(30.5) = 3.219$, $p = .003$) and probabilistic nonadjacent dependencies ($M = .668$, $SD = .214$, independent samples *t*-test not assuming equal variances: $t(33.5) = 3.230$, $p = .003$).

To test whether the addition of phonological cues particularly enhanced statistical learning of probabilistic nonadjacent dependencies, we contrasted the results of the current experiment with those of Experiment 1. Specifically, we compared mean performance and the number of participants who learned at least one dependency across conditions (no cue vs. phonological cue) and types of nonadjacent dependencies (deterministic vs. probabilistic). An ANOVA on the proportion of correct grammaticality judgments showed a marginally significant interaction ($F(1,84) = 3.048$, $p = .085$) between condition and type of nonadjacent dependency. Moreover, there was an interactional pattern for the number of participants who learned at least one dependency: an overall χ^2 test indicated that this number was not equally distributed over conditions and types of dependencies ($\chi^2(3) = 13.805$, $p = .003$). There

was a significant difference between deterministic ($n = 15$) and probabilistic ($n = 3$) nonadjacent dependencies in the no cue condition ($\chi^2(1) = 13.538$, $p < .001$), but there was no difference between deterministic ($n = 9$) and probabilistic ($n = 8$) nonadjacent dependencies in the phonological cue condition ($\chi^2(1) < 1$). For probabilistic nonadjacent dependencies, there was a marginally significant increase in the number of participants who learned at least one dependency ($\chi^2(1) = 3.030$, $p = .082$). For deterministic nonadjacent dependencies, there was a marginally significant decrease in the number of participants who learned at least one dependency ($\chi^2(1) = 3.300$, $p = .069$).

The presence of phonological cues enabled 5 more participants (22.7%) to learn at least one probabilistic nonadjacent dependency in the current experiment. The number of participants who learned at least one deterministic nonadjacent dependency was lower than in Experiment 1, but there was no significant decrease in mean performance ($t(42) < 1$, 95% CI difference = $-.094–.191$). Thus, the pattern of results indicates that the correct responses were more evenly distributed across multiple deterministic nonadjacent dependencies in the phonological cue condition. The distribution of correct responses over multiple dependencies in the phonological cue condition was not due to a change in response bias. An ANOVA on the number of endorsed items with condition (no cue vs. phonological cue) and type of nonadjacent dependency (deterministic vs. probabilistic) as independent variables did not show any significant interaction or main effects.

To compare the effects of visual and phonological cues, an ANOVA with condition (congruent visual cue vs. phonological cue) and type of nonadjacent dependency (deterministic vs. probabilistic) as independent variables was done on the proportion of correct grammaticality judgments. The analysis did not show any significant interaction or main effects, indicating that mean performance was similar for visual and phonological cues. However, the number of participants who learned at least one dependency was not equally distributed across conditions and types of nonadjacent dependencies ($\chi^2(3) = 6.741$, $p = .081$). For both deterministic ($\chi^2(1) = 3.300$, $p = .069$) and probabilistic nonadjacent dependencies ($\chi^2(1) = 3.273$, $p = .070$), the number of participants who learned at least one dependency was marginally lower in the phonological cue condition (deterministic: $n = 9$, probabilistic: $n = 8$) than in the congruent visual cue condition (deterministic: $n = 15$, probabilistic: $n = 14$). In the context of an experiment, an additional cue within the signal itself may have been somewhat less salient to participants than an external cue. Nevertheless, the results suggest that the phonological cues enabled participants to learn probabilistic nonadjacent dependencies. However, there is a possibility that participants only learned the rhyming endings, rather than dependencies between specific non-words. This issue is addressed in Experiment 3B.

Experiment 3B

To further substantiate the hypothesis that phonological cues may facilitate the learning of probabilistic nonad-

jacent dependencies, we conducted a follow-up experiment with two specific aims. The first was to verify that participants learn nonadjacent dependencies between specific non-words. To rule out the possibility that participants in Experiment 3A simply learned to accept rhyming sequences as grammatical, the phonological cues were only presented in the training phase (thus paralleling the method in Experiment 2). In the test phase, the rhyming endings were replaced by white noise. An additional control condition was run for the probabilistic nonadjacent dependencies, as a manipulation check. In this condition, the rhyming endings were omitted entirely (cut off) to examine the possibility that participants in the masked endings condition experienced a phoneme restoration effect and based their performance on rhyming endings they restored at test. The possibility that participants based their performance on remembered endings was additionally addressed in a systematic debriefing procedure.

A second aim of this experiment was to demonstrate that participants are not simply memorizing the training sequences but instead actually learning the dependencies between the nonadjacent elements. Preliminary support for this second aim comes from Gómez (2002), who found that participants were more sensitive to nonadjacent dependencies when there were many middle elements than when there were few. If participants memorized sequences of three elements, it would have been easier to memorize the small set of sequences containing a few middle elements presented repeatedly than to memorize the large set of unique sequences containing many middle elements. In addition, previous research has demonstrated sensitivity to nonadjacent dependencies in sequences containing a middle element that had not been presented in the training phase, both for sequences of non-words (Misyak & Christiansen, 2012; Onnis et al., 2004) and for sequences of numbers (Pacton & Perruchet, 2008). We therefore included a similar generalization test in the present experiment to investigate whether or not participants would be sensitive to deterministic and probabilistic nonadjacent dependencies in sequences containing a new middle element.

Method

Participants

Sixty-nine undergraduate students from Cornell University participated in this experiment for course credits. None of them had participated in any of the previous experiments. The data from three participants were discarded because the experiment was interrupted (1) or the participant was not fluent in English (2). Of the remaining 66 participants (26 male, 40 female, 18–26 years of age), 44 participated in the masked endings condition. In this condition, 22 participants were assigned to each type of nonadjacent dependency (deterministic vs. probabilistic). The final 22 participants participated in the control condition with probabilistic nonadjacent dependencies, in which the endings were removed in the test phase.

Materials

In the training phase, the same materials were used as in Experiment 3A. In the test phase, however, the endings of the initial and final non-words (-o, -im/-in, -a and -er) were masked by white noise or cut off. The middle elements were the same as during training for the standard test blocks, while the middle elements in the generalization test block consisted of four new non-words: hox, pilk, rab and sef. Each of these non-words was used twice as a middle element: once in a sequence containing a nonadjacent dependency and once in a sequence violating a nonadjacent dependency.

Procedure

The procedure for the training phase was the same as in Experiment 3A. The procedure for the test phase was also the same as in Experiment 3A, except that a generalization test block of eight sequences was added. The order of the generalization test block and the two standard test blocks was determined randomly for each participant. Participants in the condition in which the endings were cut off during the test were asked five debriefing questions after the experiment (see Appendix). The questions probed what regularities participants had noticed and whether or not they completed the non-words in their heads during the test. Participants who claimed to complete the non-words in their heads were administered a non-word completion test. Twelve non-words from the test phase were presented through the headphones: the 2 initial and 4 final non-words and 6 foils (middle elements). Participants had to indicate for each non-word whether or not they had completed it in their heads during the test and, if so, how.

On the basis of these answers to the question probes, the responses on the grammaticality judgment test were predicted. The initial and final non-words in each item of the grammaticality judgment test were completed with the ending provided by the participant during the non-word completion test. It was assumed that participants would judge items with rhyming initial and final elements as grammatical and items with non-rhyming initial and final elements as ungrammatical. A score of 1 was assigned if the predicted response was correct. A score of 0 was assigned if the predicted response was incorrect (e.g., when the participant had completed an item with a wrong ending). A score of 0.5 was assigned if the participant provided insufficient information to predict a response (e.g., when the participant failed to complete one or more non-words). Next, the absolute differences between predicted and actual scores were computed. A one-sample t-test was done for each participant to compare the difference to zero and test whether performance could be based on knowledge of the rhyming endings only.

Results and discussion

For Experiment 3B, only the proportion of correct grammaticality judgments was analyzed. The number of participants who learned at least one dependency was not computed, because the number of test items per dependency was larger than in the other experiments. For the condition with masked endings, a one-between (deterministic vs. probabilistic nonadjacent dependencies), one-within (standard vs. generalization test items) mixed-model ANOVA on the proportion of correct grammaticality judgments showed neither a significant effect of type of test item ($F(1,42) < 1$) nor an interaction between type of test item and type of nonadjacent dependency ($F(1,42) < 1$). These results indicate that the participants' sensitivity to nonadjacent dependencies was not affected by whether or not they had heard the middle element during the training phase. This finding rules out the possibility that participants based their performance on memorization of training sequences and suggests that they had learned the nonadjacent dependencies. This result confirms previous findings by Onnis et al. (2004) and Pacton and Perruchet (2008) that participants learn nonadjacent dependencies rather than the individual sequences of three elements.

The ANOVA also failed to show a significant main effect of type of dependency ($F(1,42) < 1$), indicating that participants were as sensitive to probabilistic nonadjacent dependencies as they were to deterministic nonadjacent dependencies when the dependencies had been highlighted by a phonological cue in the training phase. The proportion of correct grammaticality judgments for the standard items for both types of dependencies is shown in Fig. 2. Performance was above chance for both deterministic nonadjacent dependencies (standard items: $M = .673$, $SD = .267$, $t(21) = 3.050$, $p = .006$; generalization items: $M = .648$, $SD = .277$, $t(21) = 2.500$, $p = .021$) and probabilistic nonadjacent dependencies (standard items: $M = .597$, $SD = .182$, $t(21) = 2.494$, $p = .021$; generalization items: $M = .597$, $SD = .196$, $t(21) = 2.306$, $p = .031$). Thus, even when the phonological cues were masked by white noise in the test phase, participants demonstrated sensitivity to nonadjacent dependencies.

To further clarify the effect of the phonological cues on statistical learning of probabilistic nonadjacent dependencies, performance on the standard items in both conditions of Experiment 3B was compared with performance on the test items of Experiment 3A. A one-way ANOVA with condition (phonological cue present at test, phonological cue masked at test, phonological cue removed at test, no training) as the independent variable and the proportion of correct grammaticality judgments as the dependent variable showed a significant main effect of condition ($F(3,84) = 3.206$, $p = .027$). Orthogonal difference contrasts showed that performance was significantly higher in the three phonological cue conditions than in the no training control condition ($t(84) = 2.817$, $p = .006$). However, there was no difference in performance between the conditions in which the phonological cue was absent at test and the condition in which it was present at test ($t(84) = 1.174$, $p = .244$). In addition, there was no difference in test performance between the condition in which the phonological cue was masked and the condition in which it was removed ($t(84) < 1$). The latter finding indicates that performance in the masked endings condition was not based on a phoneme restoration effect, as phoneme restoration may occur when a phoneme is replaced by sound but not when it is replaced by silence (Warren, 1970), as in our

condition here with omitted endings. Moreover, although masking or removing the endings implied that the stimuli in Experiment 3B were changed from training to test, which could be a disadvantage, performance in these conditions was similar to performance in the condition in which the phonological cue was present. This suggests that, rather than learning a pattern of rhyming endings, participants learned nonadjacent dependencies between specific non-words from which their sensitivity to nonadjacent dependencies between the stems could result.

However, it remains possible that performance in the conditions in which the phonological cue was absent at test was based on a different strategy. Participants could conceivably respond correctly without having learned the dependencies between the non-words if they noted that the sequences in the training phase had rhyming endings and remembered the endings during the test. The responses to the debriefing questions of participants in the condition in which the rhyming endings were removed at test were used to address this possibility. Of the 22 participants, 11 guessed that the first and the third non-word were involved in a regularity (Question 5). Five participants described the regularity by naming pairs of initial and final non-words, one participant noted that the first and third non-words had common endings and two participants named pairs of initial and final non-words as well as remarking that the pairs had rhyming endings (Questions 1 and 4). Thus, only three participants mentioned the rhyming relationship.

Sixteen participants indicated that they completed the non-words in their heads during the test (Question 3). For these participants, their responses on the grammaticality judgment test were predicted on the basis of their responses to the non-word completion test and these predicted responses were compared with their actual responses. A one-sample t-test was done on the absolute difference for each participant to test whether performance could be based on knowledge of the rhyming endings only. For 4 participants, the test could not be performed because of a lack of variance. Three participants provided insufficient information to predict the response on any item; the fourth participant demonstrated perfect knowledge. The difference between predicted and actual responses was significantly larger than 0 ($p \leqslant .003$ for all comparisons) for 11 of the 12 remaining participants.

In summary, only two participants could conceivably have based their performance on knowledge of the rhyming endings only. However, these participants had complete knowledge of the nonadjacent dependencies: they were the ones who named all the pairs of initial and final non-words as well as their rhyming relationship and they scored 100% and 96% correct on the grammaticality judgment test. It is unclear whether they inferred their knowledge of the pairs from their grasp of the rhyming endings or vice versa, but they learned more than just rhyming endings. In all, the responses to the debriefing questions indicate that participants learned dependencies between specific initial and final elements and not just the rhyming endings.

More generally, Experiments 2 and 3 indicated that statistical learning of probabilistic nonadjacent dependencies

cannot only be enhanced by complementing the artificial language with external cues, but also by enriching it with phonological cues that make the dependent non-words more similar. This finding adds to the literature indicating that phonological regularities are useful sources of information in language learning (e.g., Brooks et al., 1993; Monaghan et al., 2005, 2007; Onnis et al., 2005; Pérez-Pereira, 1991).

General discussion

Several prior studies have identified conditions under which deterministic dependencies between nonadjacent elements in a sequence can be acquired by statistical learning (Bonatti et al., 2005; Creel et al., 2004; Endress, 2010; Gómez, 2002; Newport & Aslin, 2004; Onnis et al., 2003, 2004; Pacton & Perruchet, 2008). The present study has broadened the scope of this literature by investigating whether statistical learning of probabilistic nonadjacent dependencies is possible for the kind of one-to-many dependency relationships typical of nonadjacent dependencies in natural language.

The results of Experiment 1 suggest that probabilistic nonadjacent dependencies may be more difficult to learn than deterministic nonadjacent dependencies. Participants who were presented with sequences of three non-words containing a dependency between the first and the third non-word typically failed to learn the probabilistic nonadjacent dependencies when there were no additional cues available.[3] The two subsequent experiments demonstrated that more participants could learn to discriminate between grammatical sequences and sequences that violated a probabilistic nonadjacent dependency when they were presented with additional cues during the training phase. Experiment 2 showed that statistical learning of probabilistic nonadjacent dependencies occurred when the sequences of non-words were systematically presented together with visual referents. Experiment 3 showed that statistical learning of probabilistic nonadjacent dependencies occurred when the dependent non-words had phonologically similar endings. In both experiments, learning occurred when the dependent non-words were consistently associated with the same visual or phonological cue.

Of course, there may be other ways to enhance statistical learning of nonadjacent dependencies. As noted in the discussion of Experiment 1, statistical learning of probabilistic nonadjacent dependencies may benefit from a further increase in variability of the intervening material. In addition, there is some evidence that statistical learning of

---

[3] This finding would seem to differ from those of Remillard (2008) and Kuhn and Dienes (2005). One explanation of these differences may be found in the use of different measures of knowledge of non-adjacent dependencies (i.e., reaction times and liking judgments vs grammaticality judgments). Moreover, as training in the Remillard (2008) study was spread across several days, sleep-related effects of memory consolidation may help explain the results—especially as an overnight break between training sessions has been shown to improve the learning of nonadjacent dependencies (Vuong, Meyer, & Christiansen, 2011). In addition, the tone sequence used in the Kuhn and Dienes (2005) study incorporated a duration cue that likely facilitated the learning of the nonadjacent relationship that was the focus of their experiments.

probabilistic nonadjacent dependencies can be facilitated by prior exposure to adjacent dependencies with the same statistical structure. Lany, Gómez, and Gerken (2007) first exposed participants to non-words from two categories, each of which was preceded by two possible markers. These participants were later able to learn nonadjacent dependencies between markers and category-words in a different vocabulary, when the dependent non-words were separated by one of three intervening non-words. Other participants who had not been exposed to dependencies between markers and category words in isolation failed to learn the nonadjacent dependencies.

It is also possible that statistical learning of probabilistic nonadjacent dependencies may be facilitated by extended exposure to the structure. With regard to deterministic nonadjacent dependencies, there is mixed evidence on the effects of extended exposure. Newport and Aslin (2004) noted that increasing exposure from 72 to 720 repetitions (in 10 sessions) did not enhance statistical learning of nonadjacent dependencies between syllables in a continuous speech stream. Dienes and Longuet-Higgins (2004) investigated the learning of non-local dependencies in tone sequences of which the second half was a transformation of the first. They found that naïve participants still failed to learn this structure when exposure was increased from one to five sessions of 40 training sequences. However, participants who had prior experience with serialist music performed above chance after one session, suggesting that prolonged exposure can be beneficial. Similarly, Uddén et al. (2009) stressed the importance of extended exposure for learning complex structures. They found substantial learning of deterministic nonadjacent dependencies in letter strings after nine sessions spread over several days, each with 100 training sequences (see also Vuong et al., 2011, for potential effects of sleep on the learning of deterministic nonadjacent dependencies).

Although previous studies have indicated that staged learning (e.g., Lany et al., 2007), increased exposure (e.g., Uddén et al., 2009), or sleep-related memory consolidation (e.g., Vuong et al., 2011) may improve the learning of deterministic nonadjacent dependencies, it is currently unclear how much this would facilitate the learning of probabilistic nonadjacent dependencies. In contrast, our findings indicate that adding either a single visual or phonological cue does facilitate the learning of probabilistic nonadjacent dependencies within a single learning session and without staged input. More generally, the present study adds to a growing literature suggesting that statistical learning of nonadjacent dependencies occurs optimally when there are additional cues to the underlying structure (Creel et al., 2004; Gómez, 2002; Newport & Aslin, 2004; Onnis et al., 2003; Pacton & Perruchet, 2008).

It may be objected, though, that even if natural languages provide cues that correlate with (probabilistic) nonadjacent dependencies, these cues will arguably be less reliable than the visual and phonological cues that accompanied each sequence of non-words in Experiments 2 and 3 of the present study. Thus, although there may be multiple cues available, the combination of different kinds of partially reliable information would only result in unreliable learning outcomes. Or, as Pinker intuits (1984,

p. 49), "... in most distributional learning procedures there are vast numbers of properties that a learner could record, and since the child is looking for correlations among these properties, he or she faces a combinatorial explosion of possibilities. ... Adding semantic and inflectional information to the space of possibilities only makes the explosion more explosive." Whether statistical learning of (probabilistic) nonadjacent dependencies is subject to such combinatorial explosion is an important empirical question for future research. However, existing mathematical and computational work on the integration of multiple probabilistic information sources suggests that Pinker's intuition is incorrect.

If we assume, as a first approximation, that a learning process starts with a given hypothesis space defining possible solutions to a learning task, then the Vapnik–Chervonenkis (VC) dimension establishes an upper bound for the number of examples needed by a learning process to find a solution. Mathematical analyses of neural network learning have shown that the integration of multiple correlated cues will not lead to a combinatorial explosion but instead to improved learning (Abu-Mostafa, 1993). In this framework, the addition of a cue may result in the reduction of the VC dimension by weeding out bad hypotheses and in this way reduce the number of examples needed to learn the solution. This holds even if one or more of the cues are uncorrelated or otherwise uninformative with regard to the learning task, in which case they have no negative impact on learning. Artificial neural network simulations have further corroborated these mathematical results in the context of learning logical functions, such as XOR (Allen & Christiansen, 1996), and in natural language learning contexts involving small artificial corpora (Christiansen & Dale, 2001) and full-blown child-directed speech (Reali, Christiansen, & Monaghan, 2003).

Returning to our findings, we can place them in the context of multiple-cue integration. In Experiment 1, the difference in strength of adjacent and nonadjacent dependencies was a sufficiently strong cue to constrain the learner's hypothesis space for deterministic dependencies (adjacent transitional probability: .04, nonadjacent TP: 1), but not for probabilistic dependencies (adjacent TP: .04, nonadjacent TP: .5). Integrating the additional cues provided in Experiments 2 and 3 further constrained the learner's hypothesis space to nonadjacent structure for participants presented with probabilistic nonadjacent dependencies, enabling more participants to learn these dependencies.

In conclusion, the present study provided evidence that statistical learning of probabilistic nonadjacent dependencies is possible to the same degree as for deterministic nonadjacent dependencies when there are additional cues to the structure. Statistical learning was demonstrated when the dependencies in an artificial language were complemented with a congruent visual context and when the dependent non-words had phonologically similar endings. These findings are relevant to the acquisition of natural language, in which probabilistic nonadjacent dependencies can be involved in, for example, tense marking, subject–verb agreement or gender–agreement. The statistical learning of such nonadjacent dependencies from surface

regularities in the input may provide a scaffolding for the learning of more complex relationships, such as those involved in anaphoric reference. From this perspective, multiple-cue integration plays a key role in language learning and processing, allowing language to be as expressive as possible while still being learnable by domain-general learning mechanisms (see Christiansen, in press). Crucially, though, multiple-cue integration is not unique to language but is also fundamental to other processes ranging from vision (e.g., Tsutsui, Taira, & Sakata, 2005) to sensori-motor control (e.g., Green & Angelaki, 2010).

Acknowledgments

Appendix: Debriefing questions

1. Did you use any particular strategy or approach for deciding which of the test items followed the "rules" or patterns?
2. Did you notice something different about the test items compared to what you heard earlier? (If "yes," the participant had to specify what was different. If "no" [or participant's response did not mention shortened forms], the participant was informed that some of the words were shortened.)
3. During the test, did you complete the shortened words in your head, or did you focus on only the word parts that were presented to you? (If participants indicated that they completed the words in their head, the nonword completion task was administered).
4. In the second part, you had to decide whether or not sequences followed "the rules." Had you noticed any regularities in the first part? If yes, what kind of regularity?
5. Each sequence consisted of a 1st word, a 2nd word and a 3rd word. Which of these do you guess were involved in a regularity?

References

Abu-Mostafa, Y. S. (1993). Hints and the VC dimension. *Neural Computation, 5*, 278–288.

Allen, J., & Christiansen, M. H. (1996). Integrating multiple cues in word segmentation: A connectionist model using hints. In *Proceedings of the 18th annual cognitive science society conference* (pp. 370–375). Mahwah, NJ: Lawrence Erlbaum.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*, 451–459.

Brooks, P. J., Braine, M. D. S., Catalano, L., & Brody, R. E. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76–95.

Brown, R. (1973). *First language: The early stages*. Cambridge, MA: Harvard University Press.

Christiansen, M. H., & Dale, R. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In *Proceedings of the 23rd annual conference of the cognitive science society* (pp. 220–225). Mahwah, NJ: Lawrence Erlbaum.

Christiansen, M.H. (in press). Language has evolved to depend on multiple-cue integration. In R. Botha, & M. Everaert (Eds.), *The evolutionary emergence of human language*. Oxford, UK: Oxford University Press.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235–253.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 17–30.

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

Dienes, Z., & Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cognitive Science, 28*, 531–558.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation timing have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology, 64*, 1021–1040.

Endress, A. D. (2010). Learning melodies from nonadjacent tones. *Acta Psychologica, 135*, 182–190.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105*, 247–299.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*, 12203–12208.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458–467.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General, 134*, 521–537.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language, 36*, 967–997.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review, 16*, 486–490.

Gleitman, L. R., & Gleitman, H. (1992). A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition. *Current Directions in Psychological Science, 1*, 31–35.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186.

Green, A. M., & Angelaki, D. E. (2010). Multisensory integration: Resolving sensory ambiguities to build novel representations. *Current Opinion in Neurobiology, 20*, 353–360.

Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (in preparation). *Impaired statistical learning of nonadjacent dependencies in specific language impairment*. IA: University of Iowa.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of nonlocal musical rules: Implicitly learning more than chunks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1417–1432.

Lany, J., Gómez, R. L., & Gerken, L. A. (2007). The role of prior experience in language acquisition. *Psychological Science, 31*, 481–507.

Meier, R. P., & Bower, G. H. (1986). *Semantic reference and phrasal grouping in the acquisition of a miniature phrases structure language*.

Misyak, J. B., & Christiansen, M. H. (2007). Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language. In *Proceedings of the 29th annual conference of the cognitive science society* (pp. 1307–1313). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Misyak, J. B., Goldstein, M. H., & Christiansen, M. H. (in press). Statistical-sequential learning in development. In P. Rebuschat, & J. Williams (Eds.), *Statistical learning and language acquisition*. Berlin: Mouton de Gruyter.

Moeser, S. D., & Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behavior, 11*, 759–769.

Moeser, S. D., & Bregman, A. S. (1973). Imagery and language acquisition. *Journal of Verbal Learning and Verbal Behavior, 12*, 91–98.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition, 96*, 143–182.

Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. In H. Behrens (Ed.), *Trends in corpus research: Finding structure in data (TILAR series)* (pp. 139–163). Amsterdam: John Benjamins.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*, 259–305.

Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (2010). Measures of phonological typicality: Robust coherence and psychological validity. *The Mental Lexicon, 5*, 281–299.

Mori, K., & Moeser, S. D. (1983). The role of syntax markers and semantic referents in learning an artificial language. *Journal of Verbal Learning and Verbal Behavior, 22*, 701–718.

Nagata, H. (1977). The interaction between syntactic complexity and semantic reference in acquisition of a miniature artificial language. *Japanese Psychological Research, 19*, 90–96.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162.

O'Grady, W. (1997). *Syntactic development*. Chicago, IL: The University of Chicago Press.

Onnis, L., & Christiansen, M. H. (2008). Lexical categories at the edge of the word. *Cognitive Science, 32*, 184–221.

Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In *Proceedings of the 25th annual conference of the cognitive science society* (887–891). Mahwah, NJ: Lawrence Erlbaum Associates.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the 26th annual conference of the cognitive science society* (1047–1052). Mahwah, NJ: Lawrence Erlbaum Associates.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53*, 225–237.

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.

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604–607.

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language, 18*, 571–590.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Oxford: Blackwell.

Reali, F., Christiansen, M. H., & Monaghan, P. (2003). Phonological and distributional cues in syntax acquisition: Scaling up the connectionist approach to multiple-cue integration. In *Proceedings of the 25th annual conference of the cognitive science society* (pp. 970–975). Mahwah, NJ: Lawrence Erlbaum.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13*, 129–191.

Remillard, G. (2008). Implicit learning of second-, third- and fourth-order adjacent and nonadjacent sequential dependencies. *The Quarterly Journal of Experimental Psychology, 61*, 400–424.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 906–914.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110–114.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*, 105–134.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447–481.

Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds 'listen through' ambiguous adjectives in fluent speech. *Cognition, 100*, 389–433.

Tsutsui, K., Taira, M., & Sakata, H. (2005). Neural mechanisms of three-dimensional vision. *Neuroscience Research, 51*, 221–229.

Uddèn, J., Araujo, S., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2009). A matter of time: Implicit acquisition of recursive sequence structures. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st annual cognitive science society conference* (pp. 2444–2449). Austin, TX: Cognitive Science Society.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language, 27*, 71–86.

Van den Bos, E., & Poletiek, F. H. (2008). Effects of grammar complexity on artificial grammar learning. *Memory and Cognition, 36*, 1122–1131.

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.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392–393.

Yu, C. (2006). Learning syntax-semantics mappings to bootstrap word learning. In *Proceedings of the 28th annual conference of the cognitive science society* (pp. 924–929). Mahwah, NJ: Lawrence Erlbaum Associates.