The Developmental Trajectory of Nonadjacent Dependency Learning

Rebecca Gómez  
Department of Psychology  
University of Arizona

Jessica Maye  
Department of Communication Sciences and Disorders  
Northwestern University

We investigated the developmental trajectory of nonadjacent dependency learning in an artificial language. Infants were exposed to 1 of 2 artificial languages with utterances of the form [aXc or bXd] (Grammar 1) or [aXd or bXc] (Grammar 2). In both languages, the grammaticality of an utterance depended on the relation between the 1st and 3rd elements, whereas the intervening element varied freely. High variability of the middle element is known to contribute to perception of nonadjacent dependencies (Gómez, 2002), but the developmental trajectory of such learning is unknown. Experiment 1 replicated the study of Gómez with a younger age group and a more subtle variability manipulation. Twelve-month-olds failed to track nonadjacent dependencies under conditions tested here (Experiments 2a and 2b), but by 15 months, infants are beginning to track this structure (Experiment 3). Such learning has implications for understanding how infants might begin to acquire similar structure in natural language.

Much work has focused on how infants might learn adjacent dependencies in sequential structure. For instance, Kirkham, Slemmer, and Johnson (2002) showed that by 2 months of age, infants can track co-occurrence frequencies in streams of visually presented objects. By 8 months, infants can track conditional probabilities in running speech (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996) and in visually presented scenes (Fiser & Aslin, 2002). By 12 months, they

Requests for reprints should be sent to Rebecca Gómez, Department of Psychology, University of Arizona, Tucson, AZ 85721–0068. E-mail: rgomez@u.arizona.edu
track sequences of words in strings (Gómez & Gerken, 1999; Mintz, 1996; Saffran & Wilson, 2003). Thus, infants are adept at processing associative structure. Indeed, much of this work is reminiscent of themes found in early associationist approaches to language acquisition (e.g., Harris, 1955). However, criticisms of the early associationist work centered on the argument that the hierarchical structure underpinning language could not be learned from associations between adjacent elements (Chomsky, 1957).

Researchers have begun investigating how infants might learn nonadjacent sequential dependencies (Gómez, 2002; Santelmann & Jusczyk, 1998).¹ Such dependencies (also referred to as discontinuous or remotely connected dependencies) are extremely difficult to acquire (see Newport & Aslin, 2000, 2004), but play an important role in language. In English, linguistic material intervenes between morphosyntactically dependent auxiliaries and inflectional morphemes (e.g., is happily singing, has hurriedly walked) and between nouns and verbs in number and tense agreement (e.g., the birds in the trees are singing, the one who crosses the line first is the winner). As such, learning dependencies between nonadjacent words and morphemes is fundamental in acquiring the syntax of the language. This study examines when and how infants detect such structure.

Previous research with natural language shows that by 18 months infants track nonadjacent dependencies over as many as three intervening morphemes (Santelmann & Jusczyk, 1998). Infants this age were able to distinguish phrases like “is running” from “can running” and “is quickly running” from “can quickly running.” Gómez (2002) replicated the Santelmann and Jusczyk (1998) findings with 18-month-olds in an artificial language paradigm and also explored a mechanism by which infants acquire nonadjacent dependencies. She exposed 18-month-olds to one of two artificial languages: In Grammar 1 (G1), sentences followed the patterns $aXc$ or $bXd$ (e.g., pel-wadim-jic, vot-kicey-rud, vot-wadim-rud), whereas in Grammar 2 (G2), the relation between the first and third elements was reversed to $aXd$ or $bXc$, such that pel sentences ended with rud, and vot sentences ended with jic (e.g., pel-wadim-rud, vot-kicey-jic, vot-wadim-jic). Thus, in both languages, a and b elements were restricted to initial position, whereas c and d elements were restricted to final position. The two languages differed only in the relation between initial and final elements; for example, in G1, pel sentences ended with jic, whereas in G2, pel sentences ended with rud. The same set of X elements occurred in both grammars, and X elements varied freely without restriction. This resulted in identical relations between adjacent elements in the two grammars; in both G1 and G2, pel and vot could each be followed by any of the X elements, and any of the X elements could precede rud or jic. Because the two grammars are

¹Researchers have also investigated such learning with adults (see Gómez, 2002; Newport & Aslin, 2000, 2004; Onnis, Christiansen, Chater, & Gómez, 2003; Peña, Bonatti, Nespor, & Mehler, 2002; Perruchet, Tyler, Galland, & Peereman, 2004).
identical with respect to absolute position of elements and adjacent dependencies, they can only be distinguished by noting the relation between the nonadjacent first and third elements.

Furthermore, in natural language, the intervening categories are often open-class items comprising much larger sets than the function morphemes associated with nonadjacent structure. Hypothesizing that these set size differences might aid learning, Gómez (2002) manipulated the size of the set from which she drew the middle element (set size = 2, 12, or 24) while holding frequency of exposure to particular nonadjacent dependencies constant. She found that high variability led to better perception of nonadjacent dependencies even though the nonadjacent dependencies were equally frequent in small and large set-size conditions. Eighteen-month-olds were able to acquire the nonadjacent dependency when the intervening element came from a set of 24 possible words, but not when the intervening set size was smaller (2 or 12). She replicated this same finding with adult participants using a slightly more complex grammar.

At first glance, it might seem paradoxical that variability can aid learning. Indeed, by most accounts high variability should result in increased noise and thus decreased learning. However, Gómez (2002) argued that high variability in the large set-size condition acted to increase the salience of the nonadjacent elements compared to the middle element, and in this way facilitated learning. In particular, because certain kinds of structure might be perceptually salient or easier to process, infants may be biased to prefer this information. However, on the assumption that learning involves a tendency to seek out invariant structure (structure remaining constant across varying contexts; E. J. Gibson, 1969; J. J. Gibson, 1966), if the statistical probability of preferred structure decreases sufficiently, learners should begin to seek out other forms of information (Gómez, in press). Consistent with this argument, infants in Gómez (2002) appeared to be focusing on different types of dependencies as a function of their statistical properties. When the conditional probabilities of adjacent elements were relatively high (in the small set-size conditions) infants’ patterns of response indicated that they were computing the statistical relations between adjacent as opposed to nonadjacent dependencies. They were focused on local rather than remote statistical structure. However, when adjacent conditional probabilities were sufficiently low (when the set size was 24) adjacent dependencies are not reliable sources of structure, leading infants to focus instead on nonadjacent relations.

2Examples of function morphemes are prepositions, articles, auxiliary verbs, inflections, and interrogative words. Functors are frequent in language and convey important syntactic cues. In natural language, for example, auxiliaries such as is or was and inflectional endings like -ing tend to be fairly stable relative to the set of intervening verbs, a set much larger in size than the functors that occur on either side.
More important, Mintz (2002) recently proposed that the statistical properties of nonadjacent dependencies might lead learners to group the intervening elements as members of a category (see also Onnis, Monaghan, Christiansen, & Chater, 2004). Mintz (2003) analyzed what he termed frequent frames, defined as “ordered pairs of words that frequently co-occur with exactly one word position intervening” (p. 93), in a corpus analysis of child-directed speech and found that words embedded in frequent frames tend to be from the same category (e.g., noun, verb, preposition, adjective, adverb). Furthermore, the frames he identified in individual corpora do not appear to be corpus specific. Rather than being idiosyncratic to a corpus for a particular child, the same frames occur in different children’s corpora. In related research, Mintz (2002, 2004b) also showed that adults and children categorize words in artificial languages as a function of their co-occurrence patterns within frequent frames.

Mintz’s (2003) findings are important for suggesting how frequent frames might help infants begin to categorize words in speech. Categorization is an important linguistic milestone because once a word is identified by syntactic category, a child may extend it in novel syntactic contexts. Onnis, Monaghan, et al. (2004) used the example of a child learning to group words like play, drink, and eat in the context of frames like “I am-X-ing,” “don’t-X-it,” and “Let’s-X-now.” What happens when the child encounters a new word, hide, in the context of a frame such as “I am-hide-ing”? A child who has grouped play, drink, and eat should easily extend hide to a frame like “don’t-hide-it.” Indeed, this idea is supported in the empirical finding that 2-year-olds more readily acquire novel verbs when they are encountered in the context of frequent frames (Childers & Tomasello, 2001).

It is no great leap to see the relation between the frequent frames referred to by Mintz and the nonadjacent dependency learning identified by Gómez (Mintz, 2003; Onnis, Monaghan, et al., 2004). In both cases, learners have acquired dependencies between words from small categories (e.g., closed-class function morphemes) relative to the varying context created by larger categories (e.g., open-class lexical morphemes). As pointed out by Mintz (2003) and Onnis, Monaghan, et al. (2004), the Gómez (2002) studies add to the puzzle the demonstration that high variability of context may be instrumental in the identification of frequent frames. In turn, such frames may be useful for inferring syntactic categories.

The primary objective of the studies reported here was to investigate the development of nonadjacent dependency learning in an artificial language paradigm. Although 18-month-olds can track nonadjacent dependencies over up to three intervening elements in natural language, Santelmann and Jusczyk (1998) found no evidence that younger 15-month-olds were tracking such structure. Yet, infants this age do appear to be sensitive to the specific forms functors take, as well as their
order and distributions in sentences. For instance, by 11 months, infants distinguish phrases like “There was once a little kitten who was born in a dark cozy closet” from “There ki once gu little kitten who ki born in gu dark cozy closet,” in which the functors are replaced by nonwords (Shady, 1996; Shafer, Shucard, Shucard, & Gerken, 1998). By 15 months, they are more likely to segment stems of nonsense words that occur with a function morpheme during familiarization than stems that do not, suggesting that functional elements can facilitate the discovery of novel information (Mintz, 2004a). By 16 months, infants distinguish phrases like “The large cake is baking” from “Is large cake the baking” (Shady, 1996). Thus, between 11 and 18 months, infants become increasingly sensitive to functors, their distributional characteristics in speech, and the relations between them.

One explanation for increased sensitivity to relations between functional elements is that the amount of information infants can process over time (their processing window) increases with development (Newport, 1988; Santelmann & Jusczyk, 1998; see also Elman, 1993). By this view, although 15-month-olds are sensitive to function morphemes (Shady, 1996), because of their smaller processing windows they have greater difficulty tracking them over time compared to older infants. If this account is correct, infants should show emerging sensitivity to nonadjacent dependencies leading up to 18 months. We tested this hypothesis by presenting infants between 12 and 17 months of age with artificial language stimuli containing nonadjacent dependencies that were designed to be somewhat less complex than the natural language stimuli used by Santelmann and Jusczyk. If infants’ processing windows (and their ability to track information in time) increase with age, we would expect to see some sensitivity to nonadjacent structure in younger infants, albeit in a less robust form.

In addition, a question raised by the initial Gómez (2002) findings is whether there is something unique about the number 24, or whether a slightly smaller set-size condition—yet one that would still result in high variability of the middle element—might also result in nonadjacent dependency learning. Thus, in Experiment 1, we exposed 17-month-olds to strings with middle elements drawn from a set size of 18 to determine whether this amount of variability would be sufficient to facilitate nonadjacent dependency learning.

Skeptics of artificial language paradigms often question the relevance of this methodology for studying natural language acquisition. Consequently, we were eager to see how our results would map onto those reported by Santelmann and Jusczyk (1998). Although there are important differences between natural and artificial language learning, having to do with extended versus brief learning experiences and exposure to heterogeneous versus homogeneous materials, if we see similar developmental trends in the two types of studies, it will suggest that we are in fact tapping into related sensitivities (Gómez & Gerken, 1999; Morgan & Newport, 1981).
EXPERIMENT 1

Experiment 1 was conducted as a replication and extension of Experiment 2 in Gómez (2002) with slightly younger infants (17- vs. 18-month-olds). To examine the effect of increasing variability of the middle element, we presented infants with middle elements chosen from sets of 18 versus 12.

Method

Participants

Forty-eight infants were tested, 24 in each set-size condition. There were 8 boys and 16 girls in set size 18. The average age was 17 months 6 days (range = 16 months 13 days–18 months 25 days). Eleven additional infants were tested in this condition but were not included due to fussiness during familiarization ($n = 10$) or completing fewer than two G1 and two G2 test trials ($n = 1$). There were 15 boys and 9 girls in set size 12. The average age was 16 months 29 days (range = 16 months 10 days–17 months 18 days). Eighteen additional infants were tested in this condition but were not included for the following reasons: fussiness during familiarization ($n = 11$), insufficient number of test trials ($n = 6$), and chronic ear infections ($n = 1$).

Materials

Participants listened to strings of words produced by one of two artificial languages. G1 strings took the form $aXc$ and $bXd$. G2 strings took the form $aXd$ and $bXc$. Strings in both grammars contained the same vocabulary in the first, second, and third elements and the same adjacent dependencies, but differed in the dependencies between the first and third elements. A female speaker recorded sample strings. To eliminate talker-induced differences in individual strings and to ensure that there would be no idiosyncrasies as a function of different pronunciations used in the two languages, word tokens were spliced from the recorded strings and pasted together into new strings for both G1 and G2. The interword interval between the three words forming a string was 250 msec, whereas the between-string interval was 750 msec, so that the speech stream was clearly segmented into three-element strings separated by pauses. Strings were approximately 2 sec in duration. Variability was manipulated by varying the size of the pool from which we drew X elements such that $|X| = \{12 \text{ or } 18\}$.

Training stimuli. The a and b elements were instantiated as pel and vot; c and d, as rud and jic. The 18 X elements were wadim, kicey, coomo, fengle, loga, puser, gople, taspu, hiftam, deecha, vamey, skiger, benez, gensim, feenam, laeljeen, chila, and roosa. The set of 12 X elements consisted of the first 12 words in the list.
Infants exposed to set size 18 heard each of the $2 \times 18 = 36$ strings produced by their training language two times during training, for a total of 72 strings. We equated frequency of exposure to the nonadjacent dependencies across conditions. Therefore, infants exposed to set size 12 heard each of the $2 \times 12 = 24$ training strings three times to result in exposure to 72 strings. Training lasted approximately 3 min.

**Test stimuli.** Twenty-four strings in each language were used in the two test conditions (the subset of strings used for the set size 12 condition). G1 strings were presented in two separate sets in two random orders (12 G1 strings were in G1-set1 and 12 were in G1-set2; see Table 1). G2 strings were also distributed over two separate test sets. The four resulting sets (G1-set1, G1-set2, G2-set1, G2-set2) were each presented three times during the test (once in each of three test blocks), for a total of 12 test trials. Thus, in each block infants familiarized to each grammar (G1 or G2) received two test trials that conformed to their training grammar and two trials that violated it. Each test trial was approximately 30 sec in duration.

**Procedure**

Each infant was tested individually while seated on the caregiver’s lap in an enclosed booth using the head-turn preference procedure (see Kemler Nelson et al., 1995). An observer outside the test booth monitored the infant’s looking behavior using a button box connected to an Apple PowerMac. The experimental control

<table>
<thead>
<tr>
<th>Grammar 1</th>
<th>Grammar 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G1-set1</strong></td>
<td><strong>G1-set2</strong></td>
</tr>
<tr>
<td>vot kicey jic</td>
<td>pel kicey rud</td>
</tr>
<tr>
<td>pel puser rud</td>
<td>vot puser jic</td>
</tr>
<tr>
<td>vot wadim jic</td>
<td>pel wadim jic</td>
</tr>
<tr>
<td>vot taspu jic</td>
<td>pel taspu rud</td>
</tr>
<tr>
<td>pel skiger rud</td>
<td>vot skiger jic</td>
</tr>
<tr>
<td>vot gople jic</td>
<td>pel gople rud</td>
</tr>
<tr>
<td>vot hiftam jic</td>
<td>pel hiftam rud</td>
</tr>
<tr>
<td>pel fengle rud</td>
<td>vot fengle jic</td>
</tr>
<tr>
<td>vot coomo jic</td>
<td>pel coomo rud</td>
</tr>
<tr>
<td>pel loga rud</td>
<td>vot loga jic</td>
</tr>
<tr>
<td>pel deecha rud</td>
<td>vot deecha jic</td>
</tr>
</tbody>
</table>

*Note.* Twelve strings from a particular language were presented during a test trial in randomized order. The discriminations required of infants are extremely subtle.
program initiated trials and scored head-turn responses. To eliminate bias, both caregiver and observer listened to masking stimuli over headphones. During training, stimuli were presented simultaneously from two loudspeakers located on either side of the infant. The infant’s gaze was directed first toward a blinking middle light then toward one of two blinking sidelights (one below each loudspeaker). When the infant looked away from the side light for 2 sec, his or her gaze was again directed toward the middle. There was no relation between lights and sound during training.

During the test, each trial began with the light blinking at center. Once the infant fixated on the center light, the experimenter pressed a button to extinguish it. This action initiated blinking of one of the side lights (the one associated with the source of sound for that trial). When the infant turned his or her head in the direction of the side light by 30°, the test set for that trial played from the speaker located above the flashing light until the infant looked away for 2 sec (or until the trial played out after 30 sec). The observer recorded the direction of the infant’s head turns. The computer program tracked looking times, tracked the amount of time looking away from the source of sound (terminating trials after 2 consecutive sec), and controlled the randomization and presentation of stimuli.

The dependent measure was the amount of time an infant oriented toward the test stimulus. A significant difference in listening time to trained versus untrained strings averaged across trials would indicate that infants have acquired some sensitivity to the nonadjacent dependencies defined by their training language. We cannot predict the direction of preference with absolute certainty (whether toward novel or familiar structure), but orienting reliably longer to one stimulus type or the other would suggest that infants have become sensitive to the relations between nonadjacent elements.

Results and Discussion

Average listening times to trained and untrained strings are presented in Figure 1. There are many false starts in head-turn data (trials on which infants begin to look in the direction of the light, but then turn past to look at their mothers or immediately look away). Infants rarely actually fixate on the stimulus on these trials, and because they are so brief (often less than 1 sec), they tend to distort the data. In accord with the criterion for extinguishing a trial (looking away from the stimulus for 2 sec), we operationally defined false starts as initial fixations of less than 2 sec. An additional reason for adopting this criterion was that strings lasted approxi-

---

3Learning is defined operationally as discrimination of two stimulus types, measured by a significant difference in listening time to one stimulus versus the other. This measure gives no indication of the magnitude of learning, as would, for example, percentage correct.
mately 2 sec. Under the assumption that infants must register a minimum of one string in a condition to determine grammaticality, only listening times of 2 sec or greater would be useful for this discrimination. Approximately 14% of the trials (36 of 288) in set size 18 were discarded by this criterion, and about 16% (39 of 288) were discarded in set size 12.

The probability of a Type I error was set at .05 for all analyses. For effect size we report Cohen’s $d$, where effect sizes of .2, .5, and .8 are considered to be small, medium, and large, respectively (Cohen, 1988).

Preliminary analyses comparing the difference in listening time to trained and untrained strings in the two languages showed that there were no differences in performance as a function of training language (i.e., G1 vs. G2), $t ≤ 0.14, p ≥ .886, d ≤ .06$, thus the data were combined over this variable.

Mean listening times in set size 18 were 6.05 sec to the training language ($SEM = .44$) and 7.61 sec to the novel language ($SEM = .64$). A dependent $t$ test on listening times to trained versus untrained strings showed that infants listened significantly longer to strings from the novel language, $t(23) = −2.140, p = .043, d = .44$. Seventeen of 24 infants showed this pattern. As in Gómez (2002), infants exposed to set size 12 showed no discrimination, $t(23) = −0.158, p = .876, d = .03$. Mean listening times were 7.68 sec to the training language ($SEM = .82$) and 7.83 sec to the novel language ($SEM = .98$). Thirteen of 24 infants listened longer to the novel language. We next compared performance across the two conditions. An independent-sample $t$ test comparing difference scores for listening times to strings from the training language and strings from the other language failed to show greater learn-

![FIGURE 1](image-url)
ing for set size 18 versus set size 12 infants, $t(46) = 1.126$, $p = .266$, $d = .34$, suggesting that infants this age are transitional between set sizes 12 and 18. More important, we replicated the pattern of findings observed in Gómez’s (2002) earlier study that infants in the larger set-size condition showed greater learning.

These results provide additional support for the idea that learners seek out the sources of greatest statistical regularity in their input. Infants’ failure to learn nonadjacent dependencies in the smaller set size 12 condition presumably reflects their focus on the conditional probabilities between adjacent words, which are higher in this condition than in the larger set size 18 condition, but do not enable them to distinguish the two grammars. However, when adjacent probabilities were low (in the set size 18 condition), infants were able to focus on the more stable nonadjacent structure, and thus the easier learning task. The fact that infants discriminated in the larger set-size condition shows that there is nothing special about a set size of 24 (used by Gómez, 2002). Rather, what is critical is that there be sufficient variability in the middle element before learners will focus on nonadjacent dependencies in sequential structure. What constitutes “sufficient” presumably varies as a function of infant age and other factors contributing to task difficulty.

We also showed nonadjacent dependency learning in a slightly younger age group (17- vs. 18-month-olds). Infants showed discrimination after very short exposure (3 min), and the discrimination required was extremely subtle (e.g., *pel kicey rud* vs. *pel kicey jic*). Thus, the learning is impressive. The failure to discriminate in the smaller set size in combination with discrimination in the larger one suggests that infants may process adjacent dependencies as the default, only switching their focus to nonadjacent dependencies when conditional probabilities between adjacent elements are sufficiently low. Having replicated Gómez’s (2002) original findings, we next tested 12-month-olds to begin investigating the developmental trajectory of this learning.

**EXPERIMENT 2**

To establish a lower limit on nonadjacent dependency learning, in Experiment 2 we tested infants several months younger (12 months). Given their success at 12 months at tracking adjacent dependencies in artificial grammars, where infants readily learn associations between words in sentences (Gómez & Gerken, 1999; Saffran & Wilson, 2003), it seems reasonable to expect that 12-month-olds might also be able to track nonadjacent structure. Although Santelmann and Jusczyk (1998) failed to find sensitivity to nonadjacent dependencies in infants younger than 18 months, we wondered whether 12-month-olds would be successful with simpler artificial language materials. The natural language materials
used by Santelmann and Jusczyk were potentially more difficult because the nonadjacent dependencies were embedded in relatively complex passages.

**EXPERIMENT 2A**

**Method**

**Participants**

Twenty-four infants were tested (16 boys and 8 girls) with an average age of 12 months 10 days (range = 11 months 15 days–12 months 27 days). Eleven additional infants were tested but not included for the following reasons: fussiness during familiarization ($n = 4$), insufficient number of test trials ($n = 4$), technical problem running the procedure ($n = 1$), gestational term less than 37 weeks ($n = 1$), and stroke at birth ($n = 1$). Half of the infants were exposed to G1, and the other half were exposed to G2 during familiarization.

**Materials and Procedure**

The stimuli and procedure were the same as in Experiment 1, with the exception that infants were run only in the set size 18 condition.

**Results and Discussion**

As in Experiment 1, we discarded trials less than 2 sec in duration. By this criterion, 9.58% (or 23 of 288) trials were discarded. Preliminary analyses comparing listening time differences to strings from the training language versus the other language showed that there was no effect of training language (G1 or G2); thus, the data were combined over this variable, $t(22) = -0.381$, $p = .707$, $d = .15$. In the comparison of listening times to trained and untrained strings, mean listening times were 6.66 sec to the training language ($SEM = .53$) and 7.64 sec to the other language ($SEM = .61$), but this comparison failed to reach significance, $t(23) = -1.449$, $p = .161$, $d = .30$. Only 14 of 24 infants listened longer to strings from the novel language versus the training language.

These results suggest that 12-month-olds are not yet able to track nonadjacent dependencies in the materials used here, or at the most are only beginning to do so. If 12-month-olds are in a transitional period with respect to tracking nonadjacent dependencies, perhaps a larger intervening set size would enable them to focus on these relations. To investigate this possibility, in Experiment 2b we presented 12-month-olds with an intervening set size of 24.
EXPERIMENT 2B

Method

Instead of using the materials from Experiments 1 and 2a, we used those from Gómez (2002). These materials differed in the number of X elements (there were 24) and in the number of strings used during the test.

Participants

We tested twenty-four 12-month-old infants (12 boys and 12 girls), with an average age of 12 months 16 days (range = 11 months 18 days–13 months 17 days). Twelve additional infants were tested but were not included for the following reasons: fussiness during familiarization (n = 3), gestational term less than 37 weeks (n = 3), birth weight less than 5 lb 8 oz (2,495 g; n = 5), and ear infection at time of testing (n = 1).

Materials

The two languages were identical to those used in Experiments 1 and 2a such that the a and b elements were instantiated as pel and vot, and c and d were instantiated as rud and jic. In addition to the 18 X elements used in Experiments 1 and 2a (wadim, kicey, coomo, fengle, loga, puser, gople, taspu, hiftam, deecha, vamey, skiger, benez, gensim, feenam, laeljeen, chila, and roosa), we added plizet, balip, malsig, suleb, nilbo, and wiffle for a total of 24 X elements. Infants heard each of the 2 × 24 = 48 strings produced by their training language one time during training. Instead of a gap of 750 msec between strings there was a gap of 1,000 msec. As in the previous experiments, strings were 2 sec in duration. Training lasted approximately 2.5 min.

Because Gómez (2002, Experiment 2) tested a lower set size of 3 with two nonadjacent dependencies, there were only six test strings associated with each language in the set size 3 condition. These were used as test stimuli in the three set-size conditions (3, 12, and 24) in that study. The same strings were used as test stimuli in this study (see Table 2). G1 and G2 strings were presented in separate sets in each of two random orders. The four resulting sets (G1-order1, G1-order2, G2-order1, G2-order2) were each presented twice during the test (once in each of two test blocks), for a total of eight test trials. Each test trial was 17 sec in duration.

Procedure

The procedure was identical to that used in Experiments 1 and 2a.
Results and Discussion

Approximately 12% of the trials (23 of 192) were discarded for being less than 2 sec in duration. Preliminary analyses showed that there were no differences in listening time to trained and untrained strings as a function of training language (G1 or G2); thus, data were combined over this variable, \( t(22) = -0.063, p = .950, d = .02 \). Mean listening times were 6.86 sec to the training language (SEM = .49) and 6.49 sec to the other language (SEM = .50). A dependent \( t \) test comparing listening times to trained and untrained strings showed no discrimination, \( t(23) = 0.523, p = .606, d = .11 \). Fifteen of the 24 infants listened longer to the training language. Thus, even with greater variability in the middle element, 12-month-olds showed no evidence of tracking nonadjacent dependencies in sequential structure, at least under our specific testing conditions.

The a, b, c, and d elements in our artificial languages mimic function morphemes in that they consist of a much smaller set of items than the set of X elements. Whereas 12-month-olds are clearly beginning to show sensitivity to function morphemes in natural language (Shady, 1996; Shafer et al., 1998), Experiments 2a and 2b suggest that infants this age are not readily tracking the relations between functor-like elements in sequential structure, or else they may be in a transitional state with respect to this ability. These findings are consistent with those reported by Shady, who found that 10.5-month-olds distinguish phrases like “There was once a little kitten” from “There ki once gu little kitten” (in which functors are replaced by nonwords) but do not distinguish sentences like “The large cake is baking” from “Is large cake the baking” (in which real functors appear in inappropriate locations). Although these young infants are familiar with the phonetic form of functors in their language, they have not yet learned how these functors fit into sentential structure.

**TABLE 2**

<table>
<thead>
<tr>
<th>Test Strings Used With 12- and 15-Month-Olds (Experiments 2b and 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grammar 1</strong></td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>pel wadim rud</td>
</tr>
<tr>
<td>vot wadim jic</td>
</tr>
<tr>
<td>pel kicey rud</td>
</tr>
<tr>
<td>vot kicey jic</td>
</tr>
<tr>
<td>pel puser rud</td>
</tr>
<tr>
<td>vot puser jic</td>
</tr>
</tbody>
</table>

*Note.* All six strings from a particular language were presented during a test trial in one of two randomized orders.
There are at least three potential explanations for 12-month-olds’ difficulty in tracking nonadjacent dependencies. First, perhaps younger infants are less adept than older ones at noting elements that change at a slower rate (the functors) relative to those that change at a faster rate (the lexical items). This is unlikely, however, given the precision much younger infants exhibit in their sensitivity to statistical structure (Aslin et al., 1998; Fiser & Aslin, 2002; Saffran et al., 1996). Second, perhaps infants this age need greater variability in the middle element before they will begin to notice nonadjacent dependencies. This also seems unlikely, given the high variability in the set size 24 condition. Third, perhaps younger learners have a shorter processing window than older ones and thus are more restricted in their ability to process information over time (Elman, 1993; Newport, 1988; Santelmann & Jusczyk, 1998). Regardless of the reason for 12-month-olds’ failure at this task, infants’ ability to focus on nonadjacent dependencies appears to be fairly well established by 17 months. Still looking to establish a lower age limit on nonadjacent dependency tracking, in Experiment 3 we tested 15-month-old infants.

EXPERIMENT 3

The results of Experiments 2a and 2b suggest that 12-month-old infants are not tracking the particular nonadjacent relations relevant for distinguishing the two artificial languages, whereas Experiment 1 indicated that at 17 months infants are able to acquire this structure. What might we expect for intermediate-aged 15-month-olds? According to Shady (1996), by 16 months infants appear to track functors, their positions in strings, and possibly their sequential relations. These findings are at odds with those reported by Santelmann and Jusczyk (1998), who found that 15-month-olds showed no discrimination for sentences like “The boy was running” versus “The boy can running.” These inconsistencies may be indicative of a transitional state in which infants are just beginning to track the sequential structure of words in sentences. Consistent with this hypothesis, Santelmann and Jusczyk reported that a subset of their 15-month-olds (those with more advanced language abilities) were able to track nonadjacent structure. These indications that 15-month-olds are beginning to track the sequential relations between words in strings suggest that infants at this age may be at the lower age limit for tracking nonadjacent relations in our relatively simplistic artificial languages. Because we expected this task to be difficult for the infants, we presented them with a large set size of intervening elements (n = 24).

Method

Participants

We tested twenty-four 15-month-old infants (12 boys and 12 girls) with an average age of 15 months 3 days (range = 14 months 19 days–15 months 21 days).
Eleven additional infants were tested but were not included for the following reasons: fussiness during familiarization \((n = 6)\), gestational term less than 37 weeks \((n = 3)\), low birth weight \((n = 1)\), and insufficient number of test trials \((n = 1)\).

**Materials**

The stimuli were the same as those used in Experiment 2b (set size = 24).

**Procedure**

The procedure was identical to that used in Experiments 1 and 2.

**Results and Discussion**

Approximately 7% of the trials (13 of 192) were discarded for being less than 2 sec in duration. There were no differences in listening times to trained and untrained strings as a function of training language (G1 or G2); thus, data were combined over this variable, \(t(22) = 0.914, p = .37, d = .37\). Mean listening times were 8.55 sec to the training language \((SEM = .58)\) and 7.25 sec to the other language \((SEM = .42)\). This comparison resulted in a significant listening time difference, \(t(23) = 2.080, p = .049, d = .42\). Seventeen of the 24 infants showed a familiarity preference.

The fact that the 15-month-olds in this experiment produced significantly longer looks for strings that were drawn from their own training language indicates that they could discriminate the two artificial languages, and thus shows sensitivity to the nonadjacent dependencies. However, the 17-month-olds in Experiment 1 showed the opposite pattern, as did the 18-month-olds in Gómez (2002), demonstrating discrimination by looking longer for strings from the novel language. Consistent with this, a 2 (age: 17- vs. 15-month-olds) \(\times\) 2 (trained vs. untrained) mixed analysis of variance on listening time differences to trained and untrained strings resulted in a significant crossover interaction, \(F(1, 46) = 8.870, p = .005, Cohen’s f = .40\), where effect sizes of .10, .25, and .40 are taken to be small, medium, and large (Cohen, 1988).

Although both patterns of preference (whether toward familiarity or novelty) demonstrate discrimination, it is interesting that the younger infants show a reversal. Looking times have been hypothesized to be linked to the complexity of the stimuli (Hunter & Ames, 1988), with infants preferring familiar patterns when the stimuli are relatively complex, but preferring novel patterns for simpler stimuli. Because complexity is relative, the same stimuli that are somewhat complex initially may become simpler across development as infants’ processing capacity and linguistic competence increase. Previous studies have found reversals of preference that follow the predictions of this complexity hypothesis. Saffran and Thiessen (2003) found a familiarity effect for more complex stimuli but a novelty effect for stimuli that were less complex for the same age learners. Similarly,
Chambers, Onishi, and Fisher (2004) reported a switch from novelty to familiarity effects for conditions requiring infants to process simpler versus more complex phonotactic regularities. If 15-month-olds are indeed in a transitional state with respect to tracking nonadjacent structure we would expect the task in these experiments to be more difficult (or more effortful) at this younger age than at 17 months. If so, it provides a plausible explanation for the younger infants showing a familiarity effect in contrast to the older infants who showed a novelty effect.

Although the direction of preference is not critical for establishing learning (any significant difference in means indicates a sensitivity to the properties differentiating the test items), the particular preference reversal observed here makes sense if what is easier for older learners is more difficult for younger ones. Indeed, as previously noted, lesser complexity has been associated with novelty preferences and greater complexity with familiarity (Chambers et al., 2004; Hunter & Ames, 1988; Saffran & Thiessen, 2003), thus providing some precedent for this line of reasoning. However, regardless of the direction of preference, 15-month-olds do appear to discriminate the nonadjacent dependencies with a set size of 24.

GENERAL DISCUSSION

The primary aim of these studies was to investigate the developmental trajectory of nonadjacent dependency learning. Understanding such learning is important because key syntactic relations are separated by intervening morphemes, words, and phrases. In addition, nonadjacent dependencies may be instrumental in the acquisition of syntactic category structure (Mintz, 2002, 2003; Onnis, Monaghan, et al., 2004). Indeed, criticisms of early associationistic approaches pointed out their limits with respect to explaining how learners would track remotely connected structure (Chomsky, 1957).

It is clear that nonadjacent dependencies are extremely difficult to acquire (Newport & Aslin, 2000, 2004), possibly because learners are so adept at tracking adjacent structure. This preference for tracking adjacent elements is especially apparent when one considers the relative conditional probabilities between adjacent versus nonadjacent elements in these studies. In the set size 12 condition the probability that the initial word (vot or pel) will be followed by any particular X element is .083 (or 1 in 12). This probability decreases to .056 and .042 with set sizes of 18 and 24, respectively. However, in all conditions the probability that the initial element will be followed by a given third element (e.g., that vot will be followed by rud) is 1.0. Despite the enormous difference in the predictability of adjacent versus nonadjacent elements, in that nonadjacent elements were always perfectly predictable, learners appeared to track the less reliable adjacent probabilities in all but the highest set-size conditions (n = 18, 24). It is interesting that learners do not track nonadjacent dependencies with set size 12 but they do with set size 18, given that
the predictive probability of .083 between the first two elements is already low in set size 12 and only decreases to .056 in set size 18. It is as if learners are attracted by adjacent probabilities long past the point that such structure is useful.

Increased variability of intervening material does appear to lead learners to focus attention on nonadjacent structure (Gómez, 2002). In natural language, intervening elements are typically open-class words comprising much larger sets than the function morphemes associated with nonadjacent structure, and so the role of variability in such learning is plausible. Of note is the fact that adult and infant performance in the Gómez (2002) study did not increase proportionally with increases in variability (and the accompanying decrease in adjacent probabilities). Instead the shift from adjacent to nonadjacent dependency tracking occurred abruptly, and only after exposure to extremely high variability in the intervening element. This suggests that learners were focused on local information long past the point that such information might be useful, but with a sufficient decline in the usefulness of this information they were able to switch their focus.

Experiment 1 replicated Gómez (2002) and extended the findings by showing that there was nothing unique about a set size of 24. Seventeen-month-olds were as proficient at tracking nonadjacent dependencies when the intervening set size was only 18 as 18-month-olds were when the intervening words were drawn from a set of 24 (Gómez, 2002). Consistent with the earlier Gómez study, infants did not track nonadjacent dependencies in the lower variability condition, indicating that learners do not track nonadjacent structure unless variability of the middle element is sufficiently high. Although 12-month-olds did not appear able to track the nonadjacent dependencies tested here (Experiments 2a and 2b), by 15 months of age infants begin to show sensitivity to this structure (Experiment 3).

Why should there be age limitations on this learning? One possibility is that younger infants are unable to capitalize on the stable nonadjacent structure occurring in the context of high variability, but this is unlikely given work with younger infants who show exquisite sensitivity to statistical structure (Aslin et al., 1998; Fiser & Aslin, 2002; Saffran et al., 1996). Another possibility is that high variability distracts younger learners, resulting in no learning whatsoever. However, ongoing research with younger infants in Gómez’s lab suggests that high variability aids learning of anchor position in two-element strings (Gómez, Ohala, & Speranzo, 2005). A third possibility is consistent with the “less is more” hypothesis, the idea that developmental limitations on children’s processing might aid acquisition by focusing young learners on fine-grained aspects of structure (Newport, 1988). Relevant to the learning explored here, younger learners might have limits on the amount of information they can process in time (Elman, 1993; Santelmann & Jusczyk, 1998), and thus may be confined to processing information over shorter durations. If so, shortening the duration of the strings should make it easier for younger infants to track the nonadjacent structure. Infants might also be constrained with respect to the number of intervening elements that may occur be-
tween dependent elements. Because the intervening elements in this study were always two-syllable nonsense words, younger infants might fare better with only one syllable intervening. It will be important to tease apart the contributions of these two effects.

One caveat arises from the choice of stimuli used in these artificial languages. In particular, all X elements were two-syllable words with initial stress, forming a trochaic stress pattern that infants find especially attractive (Allen & Hawkins, 1978). An alternate explanation for the developmental change in performance is that younger infants were simply distracted by the appealing trochaic item, which may have diverted attention away from the critical first and third elements. In other words, it may be the case that younger infants are as capable as older ones at tracking nonadjacencies, but only older infants are able to override their trochaic bias. In future research we will need to vary the prosodic structure of the elements to address this issue. If, for example, the developmental pattern we report still obtains with strings composed of trochaic initial and final elements (e.g., kicey rud wadim, kicey pel wadim, puser vot fengle, puser rud fengle) it would rule out this potential alternate explanation for these findings.

Comparisons With Other Studies

Why do infants in our studies succeed in learning nonadjacent dependencies between arbitrary words when adults in Newport and Aslin (2004) failed to learn nonadjacent dependencies between arbitrary syllables? One key difference is that learners in their task had the added difficulty of segmenting words from continuous streams, whereas our strings were already segmented. This alone makes the task required of learners very different. However, when all other aspects of the two tasks are equated, the results are the same: Like Newport and Aslin (Experiment 1), Gómez (2002) failed to find learning in her set size 3 condition. Another key difference was the variability manipulation in Gómez. Newport and Aslin subsequently found nonadjacency learning when they replaced arbitrary syllables with segments of the same type (e.g., vowels) separated by elements of a different type (e.g., consonants), a condition that should make the task of learning nonadjacent dependencies easier (Newport & Aslin, 2004, Experiments 2 and 3). Unlike learners in Newport and Aslin, ours have no a priori reason to treat remotely connected elements as the same type. In our case, however, high variability enabled the infants to detect nonadjacent relations.

Our results might also appear discrepant with those reported by Peña, Bonatti, Nespor, and Mehler (2002), who found that learners could acquire nonadjacent dependencies with low variability in the middle element. However, the Peña et al. stimuli contained cues to grouping that may have inadvertently led to learning. In

---

We thank an anonymous reviewer for bringing this alternate explanation to our attention.
particular, their materials contained stop consonants in the first and third positions, whereas the middle elements were of a different type (liquids and fricatives). Newport and Aslin (2004) found no evidence of nonadjacent learning when all syllables contained stop consonants and suggested that such learning is facilitated by low-level grouping cues. Indeed Onnis, Monaghan, Richmond, and Chater (in press) were only able to replicate Peña et al. when grouping cues were present, confirming this hypothesis.

In addition, why do 12-month-olds fail to track nonadjacent structure when 7-month-olds in previous reported studies appear able to do so (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; see Gómez & Gerken, 1999, Experiment 4, for similar but more complex learning with 12-month-olds)? We would argue that Marcus et al. (and Gómez & Gerken, 1999) were tapping into a different learning mechanism, namely sensitivity to repetition structure (Gómez & Gerken, 2000; Gómez, Gerken, & Schvaneveldt, 2000). Identity is highly salient for learners. According to the Gestalt principle of similarity, items that are physically similar tend to be grouped together. The perceptual tendency to group by similarity could have resulted in very different learning than that observed here. In discriminating ABB versus ABA patterns for instance, infants could have generalized these as “close” versus “far” repetitions. When discriminating AAB versus ABB patterns, infants could have encoded these as repetitions occurring at string beginning or ending.

Mechanisms of Learning

The findings reported here provoke interesting theoretical questions and pose avenues for future empirical work. One question has to do with the statistic underlying learning. Gómez (2002, in press) suggested that learners attend to adjacent dependencies under small set-size conditions when dependencies between adjacent elements are still informative, but switch their focus to nonadjacent structure under conditions of high variability when conditional probabilities between adjacent elements are low. Another possible statistic is the ratio of the frequency of occurrence of beginning and end elements to middle ones. By this view, learners keep track of frequencies of elements in particular positions, but have memory limitations on the number of elements they can count across utterances. With a small number of elements, all the element frequencies can be encoded with sufficient fidelity that adjacent statistics are computed. However, as the number of elements increases, only the high-frequency elements can be encoded (and the others are treated as noise), leading learners to extract dependencies between these elements, which in these studies happen to be nonadjacent. In contrast with our original account suggesting that learners attend to adjacent dependencies as a default and that this affects

---

5We thank Richard Aslin for suggesting this statistic.
whether they encode adjacent or nonadjacent dependencies, this latter account implies that memory limitations on frequencies of individual elements affect encoding. Although the statistic on these two accounts may differ, the importance of high variability for focusing learners on stable structure is unchanged.

Another question provoked by this line of research is whether learning for both adjacent and nonadjacent dependencies is similar in nature, or as Peña et al. (2002) proposed, whether the learning involves fundamentally different mechanisms. Onnis, Destrebecqz, Christiansen, Chater, and Cleeremans (2004) provided evidence suggesting that a single associative mechanism will do. Using a simple recurrent network (Elman, 1990), they show an increase in sensitivity to nonadjacent dependencies with an increase in the variability of intervening elements. One caveat is that the Onnis, Destrebecqz, et al. network shows a gradually increasing sensitivity to nonadjacent dependencies with increasing set size. Rather than gradually increasing sensitivity, Gómez (2002) reported performance close to chance for smaller set sizes, with an abrupt increase in performance for the highest variability condition, suggesting a qualitative change in performance with increasing set size. What seems to be missing from the model is learners’ apparent focus on one source of information to the exclusion of another. It may be that Onnis, Destrebecqz, et al. are correct in proposing a single mechanism, but until the model can mimic behavior more closely, the possibility of two mechanisms is viable.

The findings reported here are consistent with a proposed mechanism for how learners might choose among multiple kinds of structure. Gómez (in press) argued that infants may be biased to prefer certain kinds of structure that are particularly salient or easy to process. On the assumption that learning involves a tendency to seek out the most stable (or invariant) forms of structure in given contexts (E. J. Gibson, 1969; J. J. Gibson, 1966), learners would need good evidence to switch to another form of structure. Such evidence could come in the form of low statistical probabilities for the preferred structure. The findings reported here were consistent with this hypothesis. In particular, infants in Experiment 1 appeared to focus on different types of dependencies as a function of their statistical properties. When local probabilities were relatively high (in the small set-size condition) infants’ pattern of response indicated that they were focused on these rather than the remote patterns reflected in the nonadjacent structure. However, when local probabilities were low (when the set size was 18) adjacent dependencies were no longer reliable sources of structure, leading infants to focus instead on nonadjacent relations. Although the evidence for the interpretation that learners will track local dependencies under low variability conditions is indirect in these studies, work in progress provides direct support for this interpretation (see Gómez, in press).

The current findings underscore the utility of artificial language approaches for understanding natural language acquisition. Such utility is sometimes questioned, due to the fact that although artificial languages provide experimental control not present in natural language, they are highly simplified. Although there are impor-
tant differences between natural language learning, which involves extended, heterogeneous exposure, and artificial language learning, which occurs with brief, homogenous exposure, it is informative to find developmental parallels in the two (Gómez & Gerken, 1999; Morgan & Newport, 1981). The developmental trajectory observed in these studies is consistent with findings emerging using natural language (e.g., Santelmann & Jusczyk, 1998; Shady, 1996). In addition, the use of artificial languages provides a level of control that permits a more delicate examination of issues for which natural language studies provide unclear results. The highly controlled patterns used in these studies enable us to account for the seemingly conflicting findings regarding 15- to 16-month-olds’ sensitivity to non-adjacent structure in natural language (Santelmann & Jusczyk, 1998; Shady, 1996). In particular, this study suggests that previous failures to find consistent behavior at this age are due to the fact that at 15 months the ability to track nonadjacent structure is emerging, but not yet robust.

The findings reported here suggest several avenues for future empirical work. For instance, a stronger test of nonadjacent dependency learning would be to present infants with strings that conform to the training dependency but contain novel middle elements. With novel middle elements there is no way that memory for specific strings can contribute to discrimination of strings that conform to versus violate the training pattern. Such discrimination is essential if nonadjacent dependency learning is to be instrumental in frame-based categorization proposed by Mintz (2003). Preliminary data in Gómez’s lab show that infants are capable of this discrimination (Gómez & Maye, 2004). Another question is at what point infants become able to process dependencies across distances greater than one intervening element. This question is particularly relevant for languages such as German in which dependencies between functional elements are separated by large distances (Santelmann, 2003). In investigating this question, Höhle, Schmitz, Santelmann, and Weissenborn (in press) discovered that the particular internal structure of the material intervening in remote dependencies can either aid or hinder learning. A final important issue is that nonadjacent dependencies in natural language can occur over abstract categories. For example, in verb agreement, it is thought to be the plural subject that controls agreement rather than the overt -s marking the noun. Thus, in future work, it will be important to begin examining nonadjacent dependencies of a more abstract kind. Studies investigating these issues are currently underway, promising additional insights into our understanding of how infants decode the complexities of language structure.

ACKNOWLEDGMENTS

This research was funded by National Institute of Child Health and Human Development Grant HD42170–01 and by National Science Foundation Grant SES–
9910203 awarded to Rebecca Gómez. We thank three anonymous reviewers for comments as well as the parents and children who participated in this research.

REFERENCES


