THE ROLE OF PRIOR EXPERIENCE IN LANGUAGE ACQUISITION

by

Jill Lany

_____

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2007

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the
dissertation
prepared by Jill Lany
entitled "The Role of Prior Experience in Language Acquisition"
and recommend that it be accepted as fulfilling the dissertation requirement for the
Degree of Doctor of Philosophy.

_____ Date: 11/29/06
Rebecca Gomez

_____ Date: 11/29/06
Merrill Garrett

_____ Date: 11/29/06
LouAnn Gerken

_____ Date: 11/29/06
Mary Peterson

Final approval and acceptance of this dissertation is contingent upon the candidate's
submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and
recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 11/29/06
Dissertation Director: Rebecca Gomez

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Jill Lany

TABLE OF CONTENTS

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

ABSTRACT

Learners are exquisitely attuned to statistical information in their language input. We tested how prior experience impacts such sensitivity, particularly whether prior experience serves as a bootstrap by enabling acquisition of more complex structure. Experiments 1 and 2 tested whether giving adult learners experience with adjacent category-dependencies in an artificial language facilitates subsequent learning of a novel language containing more complex nonadjacent dependencies. Prior experience had a facilitating effect, both when it preceded exposure to the nonadjacent language by just a few minutes (Experiment 1), and also by 24 hours (Experiment 2). Prior experience with the vocabulary and prosodic characteristics of the language did not facilitate more complex learning. Experiments 3 and 4 tested whether infants also benefit from prior experience in learning nonadjacent dependencies between categories. While 12-month-olds learn adjacent dependencies between word categories (Gómez & Lakusta, 2004), they do not track nonadjacent word dependencies until 15 months (Gómez & Maye, 2005). We asked whether experience with adjacent word-category dependencies enables 12-month-olds to generalize these relations to nonadjacent occurrences. Infants were familiarized to an artificial language containing adjacent category dependencies, and were habituated to strings in which those dependencies were nonadjacent. Infants dishabituated to strings containing violations of the nonadjacent dependencies when the dependencies had been adjacent during previous familiarization (Experiment 3), and when they were novel (Experiment 4). Infants familiarized to a language lacking co-occurrence restrictions, but otherwise matched to

the experimental language, failed to become sensitive to the nonadjacent category

dependencies during habituation. These findings demonstrate that prior experience can

bootstrap acquisition of more complex language structure.

INTRODUCTION

A promising paradigm for studying language acquisition involves using artificial grammars to investigate learning abilities (see Gómez & Gerken, 2000 for a review). Studies using this paradigm have provided evidence that learners are exquisitely sensitive to statistical information in their environment, and that such sensitivity plays a critical role in language acquisition. For example, infants can use statistical information to learn their language's phonetic categories (Maye, Werker, & Gerken, 2002), segment words from fluent speech (Aslin, Saffran, & Newport, 1996; Saffran, Aslin, & Newport, 1996; Saffran, 2001; Thiessen and Saffran, 2003), learn words' distributional characteristics (Gomez & Gerken, 1999; Saffran & Wilson, 2003), and acquire words' syntactic categories and their co-occurrence privileges (Gerken, Wilson, & Lewis, 2006; Gomez & Lakusta, 2004).

The use of artificial grammars in the above experiments allows for precise control over the cues presented to learners, and therefore more precise information about how learning occurs, which can be difficult to achieve when using natural language materials. Another advantage is that these languages are novel to learners, and therefore any learning can be attributed to controlled laboratory experience rather than prior experiences with natural language or other relevant stimuli outside of the lab. In other words, using artificial languages increases the probability that any effects are *not* due to prior experience with experiment materials.

While eliminating the effects of prior language experience is certainly an advantage in these experiments, learners acquiring their native language(s) are likely affected by

prior experience in many important ways. In particular, more complex forms of learning can be scaffolded on simpler forms (cf. Conway, Ellefson, & Christiansen, 2003; Elman, 1993; Goldowsky & Newport, 1993; Newport, 1990). Thus, to the extent that prior experience affects learning, systematically investigating the nature of such effects will shed light on an important process in language acquisition.

Recent studies of infant learning have begun to illuminate the role of prior experience in language acquisition, and on learning phonological patterns in particular. For example, Gerken (2004) and Thiessen and Saffran (2003) uncovered confounding effects of prior experience in 9-month-old infants. In both studies, English-learning infants were exposed to an artificial language in which words were characterized by stress patterns, and were then tested for discrimination between grammatical and nongrammatical strings using the Head-Turn Preference Procedure. Infants preferred test strings exhibiting a strong-weak stress pattern (the predominant stress pattern of English words) regardless of the pattern present in their familiarization language. Thus, infants' prior experience with their native language overrode the effects of the brief familiarization phase in the laboratory. Saffran and Thiessen (2003; Thiessen & Saffran 2004) tested the effects of prior experience more directly. They exposed infants to lists of words conforming to a phonological generalization (e.g. words might take the form CVCVC, or they might have a strong weak stress pattern), and then exposed them to a continuous stream of words also conforming to that generalization. Infants' prior experience with the phonological generalization allowed them to subsequently segment the fluent speech, indicating that prior experience can facilitate phonological learning.

An important point about these studies with infants is that we need not posit that prior knowledge is very abstract. For instance, strong-weak syllable sequences may share particular acoustic-phonetic manifestations that remain constant over experience. However, if prior learning is to be useful in learning syntax, a greater degree of abstraction is required. For example, there is consistency in phrase structure across phrase types. Phrases are composed of heads (the main element of a phrase, such as the noun in a noun phrase or the verb in a verb phrase) and complements (any other material in the phrase). Within a language there tends to be a consistent ordering of heads relative to complements across phrase types. For example, English is a "head-initial" language, such that the head of a phrase precedes its complement. In verb phrases, such as "walk to the store", verbs (i.e. "walk") precede complements such as prepositional phrases (i.e. "to the store"). Similarly, in prepositional phrases such as "on Jim's desk", prepositions (i.e. "on") precede complements such as noun phrases (i.e. "Jim's desk"). Thus, "walk to the store" and "on Jim's desk" have common phrase structure, although they differ greatly in their surface (acoustic) properties.

Given that phrases can differ in their surface properties, one might ask whether learners who have acquired this structure in one phrase type are able to generalize it to another. The existing infant studies do not address abstraction at this level. However, many studies of adult learning and abstraction have found that once learners are familiarized with an artificial language in one vocabulary, they can transfer to a language with the same abstract structure but novel vocabulary at test (e.g. Altmann, Dienes, & Goode, 1995; Gómez, 1997; Gómez, Gerken, & Schvaneveldt, 2000, Reber,

1969; Tunney & Altmann, 2001). In these adult studies, learners were exposed to an artificial language in one vocabulary, and were then tested on legal and illegal strings instantiated in a new vocabulary. In all of these studies, learners were able to discriminate between legal and illegal strings, suggesting that they had detected the abstract structure of the language.

Given such findings, Lany (2003) began to test how learning changes as a function of experience by giving participants multiple train-test sessions and comparing their later performance to that of naïve learners. In particular, she tested whether learners given prior experience with an artificial-language pattern will learn a new instantiation of that pattern more quickly. Participants were exposed to a language composed of words belonging to the categories *a*, *b*, *X*, and *Y*. As in natural languages, the artificial language had restrictions on how words from categories of different types could be combined within a string, such that *a* elements were paired with *X* elements and *b* elements with *Y* elements, but not vice versa. The *a*- and *b*-elements were shorter in duration and higher in frequency than the *X*- and *Y*-elements, and thus they were similar to functional elements, while *X*s and *Y*s were more like open-class elements (e.g. nouns or verbs). In this way, the *aX bY* language is analogous to the co-occurrence relationship between determiners and nouns, and also between auxiliaries and verbs in English. Thus, acquiring this artificial language recruits sensitivities relevant for learning syntactic patterns present in natural language. Figure 1 in Appendix A presents a schematic of the *aX bY* pattern.

In this study, naïve learners were given either extensive familiarization (18 minutes of exposure) or brief familiarization (6 minutes) to the *aX bY* language. A third group was given 18 minutes of exposure to the *aX bY* language, and then was transferred to a second *aX bY* language with novel vocabulary. This third group was given only 6 minutes of exposure to the second language so that it could be compared to the naïve learners given the same brief exposure. Learners were tested on their ability to discriminate between grammatical (familiar strings from training and unfamiliar strings that had been withheld from training) and nongrammatical strings from the language. Naïve learners given 18 minutes of exposure learned the language structure better than learners given 6 minutes, however neither group was able to generalize to withheld strings. In contrast, the learners given prior experience were able to generalize to unheard grammatical strings, which even learners given extensive exposure to only one language failed to do. Moreover, their discrimination for the second language was significantly better than naïve learners' even though the groups' exposure to the language was equivalent.

The findings of Lany (2003) demonstrate that learners do not remain the same over the course of acquisition, and specifically that prior learning can accelerate subsequent acquisition of similar patterns. Learners' prior experience might facilitate learning in other ways, such as increasing the robustness of learning to noise, or enabling acquisition to more difficult structure. The analogy between the *aX bY* language used in Lany (2003) and the co-occurrence relationship between determiners and nouns, and auxiliaries and verbs suggests one way that experience with simple structure might

bootstrap subsequent acquisition of more difficult structure in English: determiners and nouns are sometimes separated by an adjective (e.g. "the yellow ducky", or "a pretty flower"), and auxiliaries and verbs can likewise be separated by an adverb (e.g. "is quickly running"), or a negative element (e.g. "is not coming"). Interestingly, while languages often employ nonadjacent relationships, these can be more difficult for learners to acquire than adjacent ones (Gómez, 2002; Newport & Aslin, 2004). This raises the question of how learners become sensitive to the co-occurrence relationships in these complex constructions. One possibility is that their prior experience with adjacent co-occurrence relationships in simple constructions increases sensitivity to similar nonadjacent relationships. Speech addressed to infants and toddlers tends to consist of shorter, simpler utterances than speech to adults (Newport, Gleitman, & Gleitman, 1977; Pine, 1994), and thus learners may encounter adjacent dependencies between categories before nonadjacent ones. Thus, presenting learners with simpler instances of a pattern before more complex ones has ecological validity. Therefore, the experiments presented here tested whether experience with an *aX bY* language facilitates learning an *acX bcY* language in which these elements are nonadjacent.

Language Structure

All of the experiments presented here used a category-induction language based on Braine (1987), Frigo and MacDonald (1998), Gerken, Wilson, and Lewis (2005), Gómez and Lakusta (2004), and Lany (2003), which we refer to as an *aX bY* language. As described above, the *aX bY* language consisted of 4 word categories (*a*, *b*, *X*, and *Y*) and restrictions on how elements were to be combined into phrases. Specifically, *a*

elements were paired with *X*-elements and *b*-elements with *Y*-elements, but not vice versa.

Learners are most successful with *aX bY* languages that have relatively few *a*- and *b*-elements relative to *X*s and *Y*s (Valian & Coulson, 1988), and we therefore constructed the language so that there were three times as many *X*s and *Y*s as there were *a*s and *b*s. Also, learning this structure is facilitated when the language has cues signaling *X/Y* words' category membership (Braine, 1987; Frigo & MacDonald, 1998; Gerken et al., 2005; Gómez & LaKusta, 2004; Wilson, 2002). In these experiments, *X*- and *Y*-words were either distinguished by their syllable number, or by the presence of a distinctive ending. When syllable number was used to distinguish *X*- from *Y*-words, *X*-words were bisyllabic and *Y*-words were monosyllabic. When distinctive endings were used, *X*-words ended in the sound "-ee" or "-it" and *Y*-words ended in the sound "-oo" or "-ul".

The structure of the *aX bY* language is multidimensional (Braine, 1987; Frigo and MacDonald, 1998). One dimension concerns the predictive relationships between the *a*- and *b*-elements and the features cueing the category membership of *X*- and *Y*-elements. More specifically, each of the two *a*-elements predicts the distinctive ending present on *X*-elements, and the two *b*-elements each predict the distinctive ending on *Y*-elements. Learners sensitive to these predictive relations (or co-occurrence dependencies) can generalize to novel strings in which *a*- and *b*-elements are paired with *X*- and *Y*-elements containing the distinctive feature. These strings could contain a novel *aX* pairing (for instance in Figure 1, if a_1X_4 but not a_2X_4 was presented in familiarization), or could

involve a novel *X*-element with the distinctive feature (i.e. a_1X_7). The other dimension

of the language structure concerns the distribution of elements within strings. Because

a-elements always predict *X*-elements, *a*-elements share a distribution (similarly, *b*-

elements always predict *Y*-elements and thus also share a distribution). Moreover, the

distribution of elements in the *aX bY* language is non-overlapping, such that all *a*-

elements predict one set of elements, while all *b*-elements predict an entirely different

set. Learners sensitive to the distributional properties of elements can generalize based

on the fact that *a*-elements predict *X*- elements whereas *b*-elements predict the

nonoverlapping set of *Y*-elements. Upon hearing the string a_1X_6, they can generalize to

an a_2X_6 pairing (while rejecting a b_1X_6 or b_2X_6 pairing), even if X_6 is not marked by the

presence of the distinctive ending. Thus, these learners can generalize on the basis of

the abstract *aX bY* relation, even in the absence of the distinctive features. Previous

research done with *aX bY* languages indicates that sensitivity to this level of abstraction

is heavily dependent on the presence of correlated cues to category membership

(Braine, 1987; Frigo & MacDonald, 1998; Gerken, Wilson, & Lewis, 2005). For

example, noticing that *a*- and *b*-elements predict nonoverlapping sets of elements

requires that a majority of elements comprising those sets contain some other cue to

their category membership (i.e. the presence of distinctive endings). Therefore, the two

dimensions of structure intersect, such that sensitivity to the fact that *a*- and *b*-elements

have nonoverlapping distributions typically rests on noticing the dependencies between

*a*s and *b*s and the highly frequent endings on *X*- and *Y*-elements. These features are

sometimes referred to as "correlated cues", as they are totally or partially coupled to (or redundant with) the distributional cues to category membership.

In these experiments, we tested learners' ability to generalize to strings containing features cueing category membership. Successful discrimination of this type indicates sensitivity to the association between each of the *a*- and *b*-elements and the distinctive endings, but does not require successful sensitivity to the fact that *a*s and *b*s predict nonoverlapping sets. It does indicate abstraction beyond sensitivity to heard strings, and thus represents an important aspect of learning co-occurrence dependencies between categories. We tested this form of learning in the present experiments. Experiments 1 and 2 tested whether adults benefit from prior experience in nonadjacent dependency learning. Experiments 3 and 4 addressed this question in 12-month-old infants.

ADULT EXPERIMENTS

Experiment 1

In Experiment 1, we exposed adult learners to an *aX bY* language, and transferred them to an *acX bcY* language that was instantiated in all new vocabulary. Like the *aX bY* language, the transfer language required learning about co-occurrence restrictions on word categories. However, it differed in that the categories were separated by a *c*-element. We compared this group to a group of naïve learners exposed only to the *acX bcY* language. We predicted that learners given prior experience with the *aX bY* language would learn the nonadjacent relationships in the *acX bcY* language better than learners lacking prior experience.

Method

Participants

Seventy-nine University of Arizona undergraduates participated for course credit. Data from 14 participants were excluded for answering either "yes" or "no" to all questions in either test block (8 were from the Transfer group, and 6 from the Control group). One additional participant reported hearing loss, and those data were also excluded. Of the remaining 64 participants, 32 were in the Transfer condition.

Materials

The language materials were based on those used by Lany (2003), and consisted of two artificial languages, an *aX bY* language containing adjacent dependencies, and an *acX bcY* language containing nonadjacent ones.

The *aX bY* Language

We constructed two versions of an *aX bY* language, each with its own vocabulary (See Table 1 in Appendix A). In both versions the *a-* and *b*-elements were monosyllabic CVCs and VCs. In Version A, *a*s were alt and pel, and *b*s were erd and vot, and in Version B *a*s were ush and dak and *b*s were ong and rud. The *X*- and *Y*-elements were bisyllabic, and were distinguished from each other by the presence of a distinctive ending. For example, in Version A, all *X*s ended in the syllable "oo", and all *Y*s ended in the syllable "ee", and Version B the *X*s ended in the syllable "it" and *Y*s ended in the syllable "ul".

Each version had two grammars, *aX bY* and *aY bX*. Nongrammatical strings contained the same words and word endings as grammatical ones, but differed in the permissible combinations of word categories. For example, strings from G1 of Version A were *alt tamoo* and *erd suffee*, and strings from G2 were *alt suffee* and *erd tamoo*. In each version there were two *a*-elements and two *b*-elements, as well as six *X*-elements and six *Y*-elements. There were thus 24 legal strings in each of the grammars. In each grammar, 4 strings were withheld from familiarization to test generalization, and thus familiarization materials consisted of 20 unique strings (10 *aX* and 10 *bY* in G1, and 10 *aY* and 10 *bX* in G2.).

The test materials consisted of 16 strings, half grammatical and half nongrammatical. Four were grammatical strings that had been withheld during familiarization (grammatical-unheard strings, or GUH), and four were strings that had been presented during training (grammatical-heard strings, or GH). The

nongrammatical strings were four each of two types of strings from the other grammar.
There were NGH strings (the GH strings in the other grammar), which were "matched"
to GH strings in that they consisted of the same *X-* and *Y*-elements paired with an
ungrammatical *a-* or *b*-element. Thus, if a_1X_1, or "alt feenoo", was a GH test string in
G1, then b_1X_1, or "erd feenoo", was a NGH test string. The NGUH test strings were
nongrammatical strings matched to GUH strings, such that if a_2X_2, or "pel wifoo", was
a GUH test strings in G1, then b_2X_2, or "vot wifoo" was a NGUH test string. The test
strings are listed in Table 1 (in Appendix A).

To create the language materials, we recorded a trained female speaking the strings
in an animated voice. Because each version's two grammars differed only in the ways
that words were combined, the same token of each word was used in strings from both
grammars. This ensured that participants could not distinguish between grammatical
and ungrammatical strings based on idiosyncrasies in the pronunciation of words in one
grammar or the other. The strings were digitized and edited with SoundEdit software.
Strings were approximately 1.7 s in duration, and were separated by 1 s of silence when
presented during familiarization. Words within a string were separated by 10 ms of
silence.

The *acX bcY* Language

The *acX bcY* language was identical to the *aX bY* language, with the exception that
a *c*-element separated *a-* and *X*-elements, and *b-* and *Y*-elements (See Table 2 in
Appendix A). There were two versions of the language with distinct vocabulary
created from Versions A and B of the *aX bY* language. Each version of the *acX bcY*

language had 3 monosyllabic *c*-elements. In Version A, they were "hes", "kaf", and "sij", and in Version B they were "tash", "fis", and nep". Each version of the language had 2 grammars, such that the pairings were *acX bcY* in one grammar, and *acY* and *bcX* in the other. Thus, strings from G1 of Version A were "alt hes tamoo" and "erd hes suffee", and strings from G2 were "alt hes suffee*"* and "erd hes tamoo*"*. The set of legal strings generated by each grammar consisted of 72 strings. Twelve were withheld from familiarization to test generalization.

There was no systematic relationship between *c*-elements and the elements adjacent to them; all *c*-elements were preceded equally often by *a*- and *b*-elements, and were followed by *X*-elements as often as by *Y*-elements. Moreover, the *ac* and *bc* combinations occurred with relatively high frequency, thus the adjacent relationships were likely to be learned at the expense of the critical nonadjacent ones (see Gómez, 2002). Because adjacent relationships could not be used to discriminate between strings from the two grammars, learners focusing on them would be unable to distinguish grammatical from ungrammatical strings.

Test materials were composed of 12 strings that were withheld from training (GUH strings), 12 familiar strings (GH strings), and 12 each of the matching NGH and NGUH strings taken from the unheard grammar. These strings are listed in Table 2 (in Appendix A). The strings were divided into two sets, and were presented in two separate test blocks, each consisting of 24 strings. There were 6 each of the GH, GUH, NGH and NGUH strings in each block. Because our measures of learning assessed discrimination between grammatical strings and their nongrammatical counterparts, we

eliminated potential effects due to the possibility that some strings were endorsed at high rates because the *X*s and *Y*s were simply more memorable, easier to learn, etc. Additionally, because the *X*s and *Y*s occurring in strings withheld from familiarization occurred less frequently than the *Xs* and *Y*s occurring only in heard strings, we were able to match the frequency of occurrence of the *X*s and *Y*s in the grammatical and nongrammatical test strings.

The *acX bcY* strings were recorded by the same person who generated the *aX bY* materials and were generated in the same manner. Words within a string were separated by 10 ms of silence, and strings were separated by 1 s of silence. Strings were an average of 2 s in duration. Each block of familiarization consisted of 60 unique strings, and was approximately 3 minutes in duration.

Procedure

There were eight between-participant conditions, resulting from the manipulation of familiarization (Control vs. Transfer), language version (Version A vs. Version B), and grammar (Grammar 1 vs. Grammar 2).

Aside from instructions at the start of the familiarization phase, which were delivered by the experimenter, the entire experiment was conducted on a PC running SuperLab Pro 2.01 software. Before beginning the experiment, participants were given verbal and written instructions that they would be listening to an artificial language in which nonsense words were arranged in strings or groups. They were told to listen carefully, as they would be asked questions about the language later. Participants listened to the language and test strings over headphones.

There were two familiarization conditions (Control and Transfer). The Control participants were familiarized with 4 blocks (approximately 12 minutes) of either Version A or B of the *acX bcY* language and then tested. Transfer participants were familiarized with and tested on both the *aX bY* and the *acX bcY* category-induction languages. They were first familiarized with 18 blocks (approximately 18 minutes) of one version of the *aX bY* language and were tested, and then 4 blocks (approximately 12 minutes) of the other version of the *acX bcY* language. Thus, half of the Transfer participants were exposed to 18 blocks of Version A of the *aX bY* language, and then to 4 blocks of Version B of the *acX bcY* language (the other half were exposed to Version B followed by Version A). The learning phases were separated from each other by a brief pause in which participants were told they were done with the first phase and would next begin the second. They were then given instructions for the second phase, which informed them that they would listen to more strings from an artificial language, and that they should pay attention as they would again be tested. Thus, both Control and Transfer participants were trained and tested on 4 blocks of an *acX bcY* language, however the Transfer group was first trained and tested on 18 minutes of one of the *aX bY* languages.

Control participants were given a test directly after their familiarization with the *acX bcY* language. Transfer participants were tested at the end of each familiarization phase. The test for the *aX bY* language consisted of 2 blocks of 16 strings, each in random order. The test for the *acX bcY* language consisted of 2 blocks of 24 strings, each in random order. Before each block of test questions, participants were given both

verbal and written instructions that the strings they had heard during familiarization

adhered to a set of rules involving word order. They were also told that they were about

to hear a set of 24 strings (16 strings for the *aX bY* test), half of which followed the

rules, and half that did not. Participants were given the opportunity to ask the

experimenter for clarification of the instructions before beginning the test. Participants

heard the test strings one at a time, and made yes/no judgments on the grammaticality

of each string using the "Y" and "N" keys on their keyboard. Participants sensitive to

the *aX bY* structure should endorse grammatical test strings, including those withheld

from their familiarization, more often than nongrammatical ones.

Results and Discussion

There were no differences in learning as a function of the version or grammar to

which participants were exposed. Thus, we collapsed across these factors in all

subsequent analyses. Endorsement rates are displayed in Table 3 (in Appendix A).

We tested whether the Transfer group learned the *acX bcY* language better than

Control participants using a three-way mixed ANOVA with a between-participant

factor of familiarization condition (Control vs. Transfer) and within-participant factors

of test block (1 vs. 2), and discrimination type (GH vs. GUH). GH discrimination

reflects the ability to distinguish grammatical-heard test strings from their

nongrammatical mates (i.e. NGH strings), and is an index of sensitivity to familiar

phrases. Similarly, GUH discrimination reflects the ability to distinguish grammatical-

unheard test strings from NGUH ones, and measures generalization to novel phrases.

The dependent measure of discrimination was *d'*, and Table 4 (in Appendix A) contains

means discrimination values, both in terms of *d'* and differences in endorsement rates

for both GH and GUH strings (or "hits" minus "false alarms").

The three-way ANOVA revealed a marginal effect of familiarization condition, *F*

(1, 62) = 3.3, *p* = .074, with the Transfer group showing better discrimination (*M* = .51,

SE = .193) than the Control group (*M* = .01, *SE* = .193). There was also an effect of

discrimination type, *F* (1, 62) = 4.45, *p* = .039, with discrimination for GUH strings (*M*

= .43, *SE* = .161) better than that for GH strings (*M* = .08, *SE* = .158). Paired sample *t*

tests indicated that Control participants did not discriminate GH strings from

nongrammatical ones in either test block, *t*s (31) ≤ 1.27, *p*s ≥ .212. They did not

discriminate GUH strings from nongrammatical ones in test-block 1, *t* (31) = .08, *p* =

.931, however they did in test-block 2, *t* (31) = 2.44, *p* = .021. This finding reflects, at

most, learning during test. In contrast, Transfer participants showed discrimination for

GUH strings in both test blocks, *t*s (31) ≥ 2.13, *p*s ≤ .041, while discrimination for GH

strings was not significant in either test block, *t*s (31) ≤ .95 , *p*s ≥ .348.

Thus, Experiment 1 provides some evidence that Transfer learners' prior experience

with an *aX bY* language facilitated their learning of the more difficult *acX bcY*

language. Transfer learners' overall discrimination was marginally better than the

Control learners'. Transfer learners also showed significant discrimination for GUH but

not GH strings across both blocks of testing, while Control learners showed no reliable

evidence of learning. We attempted to provide stronger evidence that prior experience

facilitates learning nonadjacent dependencies in Experiment 2.

Experiment 2

Because Experiment 1 Transfer learners' two train-test phases lasted approximately an hour, they may have been fatigued by the end of the session. We might see clearer evidence of learning if participants were given a break between learning phases. We addressed this possibility in Experiment 2 by conducting the two train-test phases on consecutive days rather than in immediate succession. While we hypothesized that the 24-hour delay might benefit learners because of the opportunity to consolidate learning overnight with sleep, a greater delay between learning experiences could also diminish the effects of prior experience. Thus, this design permitted us to assess whether benefits from prior experience are limited to conditions of very short delay, or whether more distant prior experience also benefits learning.

Additionally, in Experiment 1, we compared transfer learners to naïve controls who lacked prior exposure to any aspect of the *aX bY* language. Reber and Perruchet (2003) demonstrated that use of naïve controls can be problematic, as they sometimes demonstrate inherent biases for some language strings over others at test. The presence of such biases makes it difficult to interpret differences between such control groups and groups familiarized with the language. Importantly, Reber & Perruchet found that familiarizing control group learners with randomized materials can ameliorate these problematic effects. Another methodological drawback to using a completely naïve control is that experimental and control learners are not equally experienced with nonspecific aspects of the language and experimental procedure. For example, it is possible that Transfer learners benefited not from exposure to the *aX bY* structure, but

from experience with superficial aspects of the language, such as the vocabulary or with the learn-test experimental procedure. Thus, Experiment 2 included a control group exposed to a language generated by the speaker of the *aX bY* language, with identical vocabulary and prosody, but which did not conform to the *aX bY* generalization in which *a*-elements predicted a set of words with a shared feature, and *b*-elements predict another set of words with a different shared feature. If these learners fail to acquire the *acX bcY* language, we will have evidence that prior exposure to the surface characteristics alone of the *aX bY* language is not sufficient to facilitate subsequent learning. Thus, Experiment 2 builds on the findings of Experiment 1 by testing the long-term effects of prior experience on learning, and by ruling out alternative explanations for observed benefits to learning.

Method

Participants

Seventy University of Arizona undergraduate students participated in exchange for course credit. The data from a total of 10 participants were excluded for the following reasons: failure to comply with instructions (N = 4, 1 from the Transfer group and 3 from the Control group), experimenter error (N = 5), and hearing loss (N = 1). Thirty were randomly assigned to each group.

Materials

The language materials consisted of the *aX bY* and *acX bcY* languages used in Experiment 1. We also recombined the *aX bY* materials to create an Uncued language, which was presented to Control participants on day 1. In Grammar 1 of the *aX bY*

language, a-elements were paired with X_{1-6} and b-elements were paired with Y_{1-6}, and

the opposite pairings held in Grammar 2. Because all Xs and Ys were marked with a

feature indicating their category membership, this language contained correlated cues to

category membership. In the Uncued language, Grammar 1 a-elements were paired

with X_{1-3} and Y_{4-6}, while b-elements were paired with Y_{1-3} and X_{4-6}. In Grammar 2, the

pairings were reversed, such that a-elements occurred with Y_{1-3} and X_{4-6}, and b-

elements occurred with X_{1-3} and Y_{4-6}. Thus, the Uncued language lacked correlated cues

to category membership.

A total of 24 strings were generated by the grammar of the Uncued language,

though 4 (one each of the types aX, aY, bX, and bY) were withheld from familiarization

to measure generalization to unheard strings. The test strings also contained 4

grammatical-heard strings (one of each of the four types), and 8 nongrammatical strings

(the matched grammatical heard and unheard strings from the other grammar). The test

materials therefore consisted of 16 unique strings. Table 5 (in Appendix A) depicts the

training and test strings from each of the versions and grammars of the Uncorrelated

language.

Importantly, all extraneous properties of the $aX\,bY$ language were retained in the

Uncued language. The Uncued and $aX\,bY$ languages contained the same vocabulary

items, presented with the same frequencies, thus the phonological characteristics of the

tokens and strings were equivalent. Moreover, both the Uncued and $aX\,bY$ languages

consisted of two-element strings in which a- and b-elements always occurred in first

position, and Xs and Ys always occurred in second position. Thus, the only difference

between the *aX bY* language and the Uncued language was that in the *aX bY* language, there were unique relationships between *a*-elements and *X*s, and between *b*-elements and *Y*s, while in the Uncued language *a*-elements predicted both *X*s and *Y*s, as did *b*-elements.

It is important to note that both the *aX bY* language and the Uncued language contained co-occurrence restrictions on how elements could be combined into strings, such that *a*s and *b*s predicted *non-overlapping sets* of elements. However, in the *aX bY* language, the co-occurrence restrictions were reliably cued by the distinctive endings on *X*s and *Y*s whereas they were not in the Uncued language (because *a*s and *b*s predicted words ending in both "ee" and "oo"). Sensitivity to the fact that *a*- and *b*-elements have nonoverlapping distributions typically relies on the presence of cues signaling the category membership of words (Braine, 1987; Frigo and MacDonald, 1998; Gerken et al., 2005). Additionally, in the absence of the distinctive cues to the category membership of *X*- and *Y*- elements, the ability to determine that a_1X_1 (if it had been withheld from familiarization) is grammatical, while the string b_1X_1 (taken from the other grammar) is not, requires that participants have an excellent memory for the a_2X_1 string. Because Control learners are not exposed to the higher-order generalization (that *a*-elements predict a set of words with a shared feature, and *b*-elements predict words with a different feature), nor do their exposure conditions facilitate sensitivity to the fact that *a*- and *b*-elements predict non-overlapping sets, it is unlikely that they will be become sensitive to the co-occurrence restrictions of their language. If sensitivity to such structural characteristics, rather than experience with the vocabulary etc., is

responsible for subsequent gains in learning, experience with the Cued language, but not the Uncued language should facilitate learning the *acX bcY* language.

Procedure

There were eight conditions, resulting from the between-participant manipulations of familiarization condition (Control vs. Transfer), language version (A vs. B), and grammar (1 vs. 2).

Participants came to the lab at the same time on two consecutive days. On the first day, participants in both groups were exposed to 18 blocks of their respective training language followed by a test. The Transfer group was exposed Version A or B of the Cued *aX bY* language, whereas the Control was exposed to the Version A or B of the Uncued language. On the second day, participants in both groups were exposed to an *acX bcY* language, with language version counterbalanced across days. The familiarization and test procedure was equivalent to that of the Transfer group in Experiment 1 in all other respects.

Results and discussion

We first tested whether the Transfer and Control groups differed in their ability to learn the *acX bcY* language on Day 2 using a three-way ANOVA with the between-participant factor of familiarization condition (Control vs. Transfer) and within-participant factors of test block (1 vs. 2) and discrimination type (GH vs. GUH). Again, *d'* was used as a dependent measure. Table 6 displays the mean endorsement rates to the three kinds of test strings for the Control and Transfer groups on Days 1 and 2.

Table 7 contains measures of their discrimination in terms of difference scores and *d'*. Both tables can be found in Appendix A.

The ANOVA yielded a marginal effect of familiarization condition, F (1, 58) = 2.73, p = .104, with the Transfer group showing better overall discrimination (M = .73, SE = .277) than the Control group (M = .08, SE = .277). There was also an interaction between familiarization condition and test block, F (1, 58) = 6.75, p = .012. We investigated this interaction by separately comparing the groups' test performance during the first and second block of testing. The Transfer group's discrimination during the first test block (M = .90, SE = .39) was significantly better than the Control's (M = -.17, SE = .16), t (58) = 2.56, p = .013. The two groups' discrimination did not differ during the second test block, t (58) = .52, p = .603, M = .57 (SE = .39) in the Transfer group and M =.34 (SE = .18) in the control.

We next tested whether either group discriminated grammatical from nongrammatical strings at above-chance rates. One-sample *t* tests indicated that Transfer participants discriminated GH and GUH strings from ungrammatical ones in the first test block, *t*s (29) ≥ 2.13, p ≤ .042, but not the second, *t*s ≤ 1.55, *p*s ≥ .13. The Control group showed no discrimination for GH or GUH strings in the first test block, *t*s ≤ -.65, *p*s ≥ .40. In test-block 2, the Control group also did not discriminate GH strings from NGH ones, t (29) = .88, p = .385, but did show significant discrimination for GUH test strings, t (29) = 2.16, p = .039. This increase is likely a reflection of learning during the test.

What can we conclude about the Transfer groups' advantage over the Control group? Our claim is that it results from differences in the two groups' prior experiences. The Transfer group was first exposed to an *aX bY* language, which consisted of two intersecting layers of structure. First, *a-* and *b*-elements predicted non-overlapping sets of elements, and second, the set of words predicted by *a*-elements shared a distinctive ending, while *b*-elements predicted a set of elements with a different distinctive ending. Learners in the Transfer group were therefore exposed to a higher-order generalization in which the co-occurrence restrictions were cued by the highly frequent endings on *X*- and *Y*-elements. Any sensitivity to such structure should benefit their learning of the *acX bcY* language. In contrast, the Control participants had prior exposure to the Uncued language, which contained co-occurrence restrictions (i.e. *a-*and *b*-elements predicted non-overlapping sets), but the restrictions were not cued by the endings on *X*- and *Y*-elements. Given that such cues are typically critical to learning (Braine, 1987; Frigo & McDonald, 1998; Gerken et al., 2005), we predicted that Control learners would be unlikely to become sensitive to these restrictions, and as a result would not benefit from their prior experience upon exposure to the *acX bcY* language.

We can test these predictions by assessing the two groups' performance on Day 1. Participants' Day 1 endorsement rates can be found in Table 6, and measures of discrimination can be found in Table 7 (both in Appendix A). Using *d'* as a dependent measure, control participants' discrimination for GH strings in test-block 1 was marginally significant, $t(29) = 1.74$, $p = .093$, and discriminations for the other strings did not differ from chance levels, ts $(29) \leq .73$, ps $\geq .457$. Thus, the Control group

showed no ability to generalize to unheard strings, and their recognition of familiar

strings was marginal. We can conclude on this basis that they were not sensitive to the

co-occurrence restrictions of their language, and even their ability to recognize familiar

strings was weak.

The Transfer group's day 1 performance was much more suggestive of learning.

Paired-sample t tests using d' as a dependent measure indicated that their Block 1 and 2

GH discriminations differed from chance, $ts \geq 2.76$, $ps \leq .011$, and they were also able

to discriminate GUH strings from NGUH ones in test-block 1, $t (29), = 3.41$, $p = .005$.

However, they did not show discrimination for GUH strings in test-block 2, $t (29) =$

$.94$, $p = .54$. This pattern of results suggests that the Transfer group showed robust

discrimination of familiar strings, and may also have developed some sensitivity to the

co-occurrence restrictions of the *aX bY* language. This sensitivity was absent in learners

exposed to the Uncued language, and is therefore likely to have promoted Transfer

learners' sensitivity to the nonadjacent relationships in the *acX bcY* language.

Confirmation for this comes from the fact that Transfer learners' sensitivity to the *aX*

bY language, as measured by their overall test discrimination, was significantly

correlated with their sensitivity to the *acX bcY* language, $r = .513$, $p = .004$, N = 30. The

Control group's performance during the two phases was not correlated, $r = -.076$, $p =$

$.689$, N = 30. One might argue that Control participants would not be expected to learn

the *aX bY* structure, and that sensitivity to GH strings is more likely to be a predictor of

subsequent sensitivity. However, it was not the case that memory for GH strings in the

in the first phase predicted their sensitivity to heard strings in the *acX bcY* language for either the Control or Transfer participants, rs < .279, ps >.135, Ns = 30.

It is possible that the Control group in Experiment 2 did not benefit because exposure to the Uncued language actually hindered subsequent learning relative to naïve learners. Mean discrimination in Control group from Experiment 2, in which learners were given prior experience with the Uncued language, was .08 (SE = .113), and discrimination in the naïve Control group was .01 (SE = .11). This difference was not significant, F (1, 60) = .22, p = .64. Thus, experience with the Uncued language did not hinder learning relative to naïve learners, and the superior performance of Transfer vs. Control participants on the *acX bcY* language in Experiment 2 cannot be due to poor performance in the Controls due to interference.

To summarize, the findings from Experiment 2 suggest that prior experience with an *aX bY* language enables learning of the more difficult *acX bcY* language. Additionally, the findings rule out the possibility that exposure to the vocabulary, prosodic characteristics, and positional regularities of *a-*, *b-*, *X-,* and *Y*-elements, rather than the *aX bY* structure itself, account for the Transfer group's ability to learn the *acX bcY* language during Phase 2. The findings also suggest that prior experience need not immediately precede exposure to a parallel structure to assist learning, as the Transfer group benefited from an experience that took place 24 hours before their exposure to the *acX bcY* language.

Discussion of Adult Experiments

Experiments 1 and 2 demonstrate that prior experience with simple structure facilitates learning more complex structure. In Experiment 1, learners with prior exposure to an *aX bY* language went on to learn an *acX bcY* category induction language, in which the co-occurrence restrictions pertained to nonadjacent word categories, marginally better than naïve participants given equivalent exposure to the *acX bcY* language. In Experiment 2 we found stronger evidence that experience with adjacent relationships facilitates learning nonadjacent ones, and that sensitivity declines over the course of testing. Additionally, we demonstrated that such experience could occur on the preceding day and still confer a benefit, and eliminated the possibility that the benefit obtained from prior experience was due to nonstructural dimensions of prior experience (i.e. vocabulary, positional information) rather than experience with the *aX bY* structure itself. The gain in sensitivity afforded by prior experience is noteworthy given that the transfer language required learning of nonadjacent co-occurrence restrictions, as opposed to adjacent ones.

Several aspects of the data in these experiments bear further discussion. First, the Transfer group from Experiment 2 benefited from prior experience with *aX bY* structure, but that learning was only expressed during the first block of testing. Why might their sensitivity have diminished over testing? One explanation has to do with the complexity of the *acX bcY* language, which contained 48 unique adjacent transitions and 24 unique nonadjacent transitions, The language potentially placed substantial demands on memory, and, as a result, sensitivity to the co-occurrence relationships may

have been susceptible to interference from nongrammatical strings heard during testing. It may be that familiarization conditions could be manipulated to produce learning more resistant to the influence of nongrammatical instances, but importantly, in real-life language learning, grammatical instances are far more common than nongrammatical ones.

Our second point concerns the Transfer learners in Experiment 1, who showed significant GUH discrimination but no GH discrimination. Given that GH strings have the extra benefit of having been heard, GH discrimination is typically better than GUH discrimination. Therefore, it is puzzling that GUH learning tended to be better. However, the fact that GH discrimination was significant in Experiment 2 suggests that the lack of GH discrimination in Experiment 1 may have been a spurious finding.

Third, these data also suggest the intriguing possibility that learners need not attain full knowledge of abstract structure during initial learning to show enhanced sensitivity to that structure in subsequent learning. Day 1 performance of Transfer learners in Experiment 2 suggests that 18 blocks of exposure is not enough to induce robust generalization to GUH strings (Transfer learners in Experiment 2 showed generalization to GUH strings from the *aX bY* language only in test-block 1 on Day 1). Thus, at the point of familiarization with the *acX bcY* language these learners did not appear to have robust knowledge of the languages co-occurrence restrictions. However, whatever sensitivity they did have did benefit them upon exposure to the second version of the language, as shown by their generalization to grammatical-unheard strings. This suggests that weak or partial learning of abstract structure can, under some

circumstances, enhance subsequent learning. By this view, exposure to two instantiations of the *aX bY* language may actually promote abstraction, perhaps more so than extensive exposure to a single version of the language. In other words, and in contrast to traditional notions of experience-driven learning (Gold, 1967), exposure to multiple instantiations of a pattern may facilitate abstraction rather than presenting learners with an intractable induction problem.

Finally, these data have implications for accounts of how learners acquiring nonadjacent dependencies. Gomez (2002) found that nonadjacent dependencies are acquired when there is substantial variability in the material that intervenes between dependent elements. One explanation for this finding is that learners attend to adjacent relationships until they become unreliable, at which point they become more attentive to other kinds of structure, such as nonadjacent dependencies (i.e. the dependencies that *a*s predict "oo" and *b*s predict "ee" in the *acX bcY* language). Another explanation is that because nonadjacent dependencies occur with the same set of intervening elements, the language contains distributional evidence that *a*s and *b*s belong to the same category. For example, in the *acX bcY* language, both *a*s and *b*s predicted the same set of *c*-elements, and thus, learners could have been encouraged to group *a*s and *b*s into one category, "*z*". If this were the case, they would fail to discriminate between *acX* and *bcX* strings not because they did not detect the nonadjacent relationships, but because they perceived them both to have the form *zcX*. When the language contains a small number of *c*-elements, there are fewer adjacent *ac* and *bc* relationships to track. To the extent that learning of adjacent relationships is better when there are fewer of

them to acquire, learners exposed to fewer *c*-elements may be more likely to form the "z" category, and therefore less likely to discriminate between *acX* and *bcX* strings, perceiving both to be of the form *zcX*.

One reason to favor the account put forward by Gomez (2002) is that while the use of the same c-elements in the *acX* and *bcY* dependencies results in distributional information that might result in a failure to discriminate these nonadjacent dependencies, learners are unlikely to use such distributional cues if they are not correlated with some other cue indicating that *a*s and *b*s belong to the same category. The language did not contain these cues, however, as *a*s and *b*s were not phonologically distinctive in the language (they had the same form as *c*-elements). Moreover, in Experiments 1 and 2 learners were exposed to an *acX bcY* language which contained a small number of *c*-elements, and only learners with prior experience with adjacent relationships were able to detect the nonadjacent dependencies. It is unclear how such prior exposure would over-ride the bias to attend to distributional cues that *a*s and *b*s belong to the same category. Rather, it is more likely that prior experience with a language containing co-occurrence restrictions on word categories promoted sensitivity to nonadjacent ones by attuning learners to the presence of relationships holding between word categories.

INFANT EXPERIMENTS

Experiments 1 and 2 provide evidence that prior experience may be instrumental to learning nonadjacent syntactic relationships, but an unanswered question is whether infants show similar benefits. As discussed previously, the speech addressed to infants and toddlers tends to consist of shorter, simpler utterances than speech addressed to adults (Newport, Gleitman, & Gleitman, 1977; Pine, 1994), and thus infants are often exposed to simple constructions before more complex ones. For example, infants are likely to hear sentences like "the ducky is in the tub", in which the predictive relationship between determiners and nouns holds between adjacent elements, before strings like "the yellow ducky is in the tub", in which these elements are separated by an adjective. Thus, it is plausible that experience with adjacent dependencies in simple constructions plays a role in infants' ability to learn nonadjacent dependencies. However, for this to be so, infants would have to generalize from adjacent to nonadjacent dependencies despite the fact that the surface features of the strings were quite dissimilar. The adult learners in Lany et al. (in press) were able do so, but there is considerable debate over the extent to which infants can achieve such generalization in early language acquisition (see Tomasello (2000) and Fisher (2002) for an interesting debate of this issue), and thus the extent to which they benefit from prior experience in language acquisition is unclear. Some data suggest that infants benefit when a new learning scenario bears a high degree of similarity to a past one (Saffran and Thiessen, 2003; Thiessen & Saffran 2004). The current study addresses whether infants can benefit from prior experience when it is not highly similar to the current learning

scenario, and specifically whether infants' experience with adjacent dependencies between word categories bootstraps sensitivity to nonadjacent ones.

We can also probe factors influencing the developmental trajectory of sensitivity to such dependencies. While infants demonstrate robust sensitivity to adjacent relationships by 8 months of age, Santelmann and Jusczyk (1998) showed that English-learning infants do not become sensitive to nonadjacent relationships in their native language until somewhere around 18 months of age. Moreover, even 18-month-olds' ability to track such relationships is not yet adult-like. While they track nonadjacent relationships spanning 3 or fewer intervening morphemes, they fail to track dependencies spanning a greater number. Based on these findings, Santelmann and Jusczyk (1998) suggested that increases in infants' working memory capacity support an enhanced ability to detect nonadjacent structure.

Gomez & Maye (2005) found further evidence of developmental differences in the ability to learn nonadjacent relationships. They found that 15- and 17-month-old infants were able to track nonadjacent relationships in an artificial language, while 12-month-olds were unable to do so. Moreover, 15-month-olds may have found the nonadjacent relationships more difficult to acquire than the older infants. Infants of both ages were familiarized to an artificial language containing nonadjacent relationships, and then were exposed to familiar strings in which the nonadjacent relationships were grammatical, and also to novel strings in which they were ungrammatical. Seventeen-month-olds showed a novelty preference, as did 18-month-old infants in a previous study by Gomez (2002); however, the 15-month-old infants showed a familiarity

preference. While direction of preference is difficult to interpret, there is evidence that infants show a novelty preference when exposed to simple stimuli, and a familiarity preference when exposed to relatively complex stimuli (Hunter & Ames, 1988; Thiessen & Saffran, 2003). Similarly, Gomez and Maye point out that infants succeed at learning nonadjacent relationships when exposed to an artificial language by 15 months, but do not demonstrate sensitivity to more complex instances taken from natural language materials until 18 months (Santelmann & Jusczyk, 1998). Together, these findings suggest that 15-month-olds were more challenged by the task than the older infants.

While maturation and concomitant increases in processing capacity surely contribute to infants' ability to learn nonadjacent relationships, prior experience may also play a critical role. Although older infants have increased processing and memory capacity, they also have more language experience. In particular, they have more accumulated experience with simpler constructions in which dependent elements are adjacent. This experience could enhance their sensitivity to nonadjacent relationships between such dependent elements, perhaps by attuning them to the presence of predictive relationships. Because Gomez and Maye (2005) found that the ability to track nonadjacent relationships does not emerge until 15 months of age, even under the most conducive circumstances, we tested whether 12-month-old infants could learn them if given relevant prior experience.

Experiments 3 and 4 therefore tested two questions about the role of prior experience; first, does prior experience with adjacent dependencies between word

categories bootstrap infants' sensitivity to nonadjacent dependencies despite the fact that it requires substantial generalization, and second, can infants younger than 15-months learn nonadjacent dependencies if given relevant prior experience? In Experiment 3, we tested whether 12-month-old infants exposed to an artificial language containing adjacent dependencies recognize the same dependencies when they occurred nonadjacently. In Experiment 4, we tested whether they generalize to novel dependencies respecting the co-occurrence restrictions present in the adjacent language. Two control groups were given prior experience with a language matched to the adjacent language in vocabulary and prosodic characteristics, but lacking the critical co-occurrence dependencies. This allowed us to determine whether exposure to the adjacent language's underlying structure is necessary for bootstrapping to the nonadjacent dependencies, or whether exposure to its surface characteristics is sufficient.

Experiment 3

Twelve-month-old infants can learn co-occurrence relationships between word categories in an artificial language (Gomez & Lakusta, 2004), while infants this age seem unable to track nonadjacent structure (Gomez & Maye, 2005). The current experiment tested whether 12-month-olds given prior experience with adjacent dependencies can recognize those dependencies when they are nonadjacent. Infants in the Experimental group were therefore familiarized to an artificial language containing adjacent dependencies, and then were habituated to strings in which those dependencies were nonadjacent. Control infants were exposed to a language lacking consistent

adjacent co-occurrence restrictions, but matched to the consistent language in vocabulary and prosody, and then were habituated to strings containing nonadjacent dependencies. Infants sensitive to the nonadjacent relationships should dishabituate to strings containing violations of those dependencies. If prior exposure to adjacent dependencies is critical to detecting nonadjacent ones, Experimental but not Control infants should dishabituate to such violations.

Method

Participants

Participants were 32 infants from the Tucson community. Infants ranged from 355 to 390 days of age (adjusted for birth term), and had an average age of 366.47 days (*SE* = 1.38). There were 16 infants assigned to the Experimental condition (7 female), and 16 were assigned to the Control condition (9 female). Within each familiarization condition, infants were randomly assigned to hear one of the two grammars of the artificial language. Infants born early (fewer than 37 weeks gestation), weighing less than 5 lbs., 8 oz. at birth, or having immediate family members with language delays were not included. Data from additional infants were excluded because of fussiness (N = 12), parental interference (N = 6), and equipment failure (N = 5).

Materials

Familiarization

Infants were exposed to an artificial language of the form *aX bY*. We used the *aX bY* language materials from Gomez and Lakusta (2004), as they demonstrated that 12-month-old infants successfully learned its co-occurrence restrictions. The language was

also similar to that used to test adults in Experiments 1 and 2, except that the feature used to distinguish *X*- and *Y*-elements was syllable number.

The language consisted of two *a*s, two *b*s, six *X*s, and six *Y*s. The *a*s and *b*s were instantiated as monosyllabic VCs and VCCs (*alt*, *ush*, *ong*, and *erd*). The feature distinguishing *X*s from *Y*s was syllable number, such that *X*s were bisyllabic CVCVCs (e.g. *coomo*, *fengle*, and *kicey*), and *Y*s were monosyllabic CVCs (e.g. *deech*, *ghope*, and *jic*). There were two grammars of the *aX bY* language, such that in G1 the pairings were *aX* and *bY* and in G2 they were *aY* and *bX*. Note that the co-occurrence restrictions permissible in one grammar were ungrammatical in the other.

In G1 there were 12 *aX* phrases and 12 *bY* phrases (similarly in G2 there were 12 *aY* phrases and 12 *bX* phrases). Phrases from the language were combined to form strings, each containing an *aX* phrase and a *bY* phrase (i.e. in G1 strings were *aX bY* or *bY aX*, and in G2 they were *aY bX* or *bX aY*). There were 288 possible strings generated by each grammar, and a subset of 48 was presented during familiarization. Each *a*- and *b*- element was presented 24 times during 1 block of familiarization, and each *X*- and *Y*- element was presented 8 times. Each of the *aX* and *bY* phrases was presented 4 times during 1 familiarization block. Table 8 (in Appendix A) depicts the *aX bY* familiarization materials.

Infants in the Control group were familiarized to an Uncued language, which was equivalent to the consistently cued *aX bY* language except that it lacked co-occurrence restrictions on word categories. The Uncued language consisted of a mixture of all of the phrases from both G1 and G2 of the *aX bY* language (see Table 8 in Appendix A).

Each phrase of the language contained an *a*, *b*, *X*, and *Y* element, however *a*s occurred

with all 6 *X*s and *Y*s, as did the *b*-elements. There were thus 48 phrases (12 *aX*, 12 *bY*,

12 aY, and 12 bX). Phrases were combined into strings containing either an *aX* and a

bY phrase, or an *aY* and *bX* phrase, such that they were identical to strings in either G1

or G2 of the *aX bY* language. Of the 576 possible strings generated by the language, a

subset of 48 were presented during familiarization. The Uncued language provided

equivalent exposure to individual vocabulary elements as the *aX bY* language (each *a*

and *b* was presented 24 times during one familiarization block, each *X* and *Y* was

presented 8 times), and each phrase was presented twice during one block of

familiarization. Additionally, the language had the same intonational patterns and

positional regularities as the *aX bY* language (i.e. *a*s and *b*s occur in first position

whereas *X*s and *Y*s occur second). Critically, however, it did not conform to an *aX bY*

generalization.

All language materials were spoken by a trained female speaker. Recorded

materials were digitized and edited using SoundEdit 16.1 software. Because the two

grammars of the *aX bY* language differed only in their permissible word combinations,

the same word tokens were used to create the strings of both grammars. The Uncued

language consisted of a combination of strings from both grammars of the *aX bY*

language. This ensured that the phonological characteristics of the two grammars of the

aX bY language and the Uncued language were matched.

Phrases in both the *aX bY* and Uncued languages had a rising intonation over the

first word, and a falling intonation over the second word, and the words within each *aX*

or *bY* phrase were separated by .03 seconds of silence. Strings, which consisted of an *aX* and a *bY* phrase, thus contained two rise-fall patterns that were separated by .4 seconds of silence. Strings were separated by 1.1 seconds of silence. In both the *aX bY* and Uncued languages, familiarization materials consisted of 48 unique strings, which were a total of about 3 minutes and 13 seconds in duration. The set of strings was randomized three times, such that infants were exposed to 3 blocks of the familiarization set.

Habituation and Test

After familiarization with strings from the *aX bY* or Uncued languages, infants were habituated to strings containing the same *a*s, *b*s, *X*s, and *Y*s as their training language, however these elements were separated by a *c* element. This allowed us to test whether familiarization to adjacent relationships facilitates detection of these relationships when they are subsequently presented nonadjacently. The habituation and test strings for Experiment 1 are listed in Table 8 (in Appendix A). The habituation materials were created by inserting a novel *c* element, *hes*, in the center of *aX* and *bY* phrases from the two grammars of the *aX bY* language. Thus, G1 habituation phrases took the form of *acX* and *bcY*, while in G2 they were *acY* and *bcX*. Both grammars consisted of 24 phrases: in G1 there were 12 *acX* and 12 *bcY* phrases, and in G2 there were12 *acY* and 12 *bcX* phrases. Within each grammar, the phrases were combined to form 24 strings: 12 *acX bcY* and 12 *bcY acX* strings in G1, and 12 *acY bcX* and 12 *bcX acY* strings in G2. Thus, as in familiarization, each phrase occurred in both the initial and the final position in the string. We created two stimulus sets composed of different

randomizations of the G1 habituation strings, and 2 sets composed of different randomizations of the G2 habituation strings. This ensured that infants were not exposed to the same order of strings on each habituation and test trial.

All strings had rising intonation over the first two words, and falling intonation over the third. As in familiarization, words within a phrase were separated by .03 seconds of silence. Strings, which consisted of an *acX* and a *bcY* phrase, contained two rise-rise-fall patterns separated by .4 seconds of silence. Strings were separated by 1.1 seconds of silence.

Infants familiarized to G1 of the *aX bY* language were habituated to G1 habituation strings, and infants familiarized to G2 of the *aX bY* language were habituated to G2 habituation strings. Infants familiarized to the Uncued language were habituated to either G1 or G2 habituation strings. After habituation, infants were tested on the habituation strings from the other grammar. The habituation and test strings were generated in the same manner as the familiarization strings; one set of word tokens was used to create both habituation and test strings, and thus pronunciation differences could not be used to distinguish between them. Moreover, habituation and test strings contained the same adjacent combinations (i.e. *ac*, *bc*, *cX*, and *cY*), and thus they could only be distinguished by their nonadjacent dependencies.

Procedure

Familiarization Phase

Infants in the Experimental group were familiarized with 3 blocks (approximately 8 minutes, 50 seconds) of either G1 or G2 of the *aX bY* language, and infants in the

Control group were given equivalent exposure to the Uncued language. Participants were thus given three times the exposure given to infants in Gomez and Lakusta (2004). We departed from their procedure to ensure that infants achieved a robust sensitivity to the $aX bY$ language, as the nonadjacent language they were subsequently exposed to during habituation was likely to be quite difficult to learn. Infants were individually exposed to their familiarization language while playing quietly with their parent(s) and an experimenter in a playroom in our laboratory. The language was played from a Sony hand-held tape-player sitting on a table in the playroom. Infants were exposed to the language in this setting to increase the likelihood that they remain comfortable during the familiarization period. Infants were allowed to play with any of the toys present in the playroom (e.g. plastic buckets, blocks, and shape-sorters, stacking cups, and stuffed animals). Parents were instructed not to talk to their infants to ensure that the familiarization language was audible, however they were otherwise allowed to interact naturally with their infants.

Habituation Phase

After infants were familiarized to the language, they were taken to an enclosed booth and seated on their parent's lap for the habituation phase. An experimenter sitting outside the booth monitored and recorded the infants' looking behavior through a closed-circuit video-recorder and TV monitor, and controlled the habituation procedure using Habit software (Cohen, Atkinson, & Chaput, 2004) running on a Macintosh computer (OS X). At the beginning of each habituation trial, a visual "attention-getter" (a colorful bull's-eye) was projected onto screen in front of the infant. Once the infant

fixated on the bull's-eye, a looping cartoon-style animation of a baby bouncing up and down in a crib replaced the bull's-eye, and habituation strings began to play from a center-mounted speaker above the screen. The image persisted, and habituation strings continued to play, until infants looked away from the screen for 2 consecutive seconds, or until the habituation trial had lasted 60 seconds. At this point the trial ended, and a new trial began. This process repeated until infants' summed listening across three trials reached a habituation criterion (50% of their listening time to the first three trials).

Test Phase

Once infants reached the habituation criterion, they were exposed to two test trials. Infants habituated to G1 (*acX bcY*) heard G2 (*acY bcX*) strings during the test trials whereas infants habituated to G2, were tested on G1 strings. Infants familiarized to the Uncued language were tested on either G1 or G2 strings. Finally infants were exposed to a "recovery trial" in which the auditory stimulus was a voice saying "boo baa" repeatedly, and the visual stimulus was, again, the bouncing baby. Given that the "boo baa" syllables were totally novel in the context of the experiment, alert infants should listen longer to these trials than to their last two habituation trials. This allowed us to exclude infants who had stopped paying attention to the auditory stimuli by the end of the experiment due to fatigue or boredom.[1]

Parents wore headphones and listened to masking music throughout the habituation and test phases to ensure that they could not systematically influence their infants'

---

[1] In Experiment 3, in was not necessary to discard any infants for this reason.

behavior. They were instructed to keep their infant comfortably seated on their lap, and to remain unresponsive to the visual stimuli presented on the screen.

Results and Discussion

We asked whether 12-month-old infants, who are typically unable to detect nonadjacent dependencies between words, will detect them if given relevant prior experience. Infants in the Experimental group were familiarized with *aX bY* strings, and habituated to strings containing familiar *aX bY* pairings separated by a *c* element. Infants in the Control group were familiarized to the Uncued language, which did not contain adjacent dependencies, but was matched to the *aX bY* language in all other respects, and were then habituated to *acX bcY* strings. If 12-month-olds cannot detect nonadjacent dependences, regardless of their prior experience, then neither group should dishabituate to test strings containing violations of nonadjacent dependencies. If experience with consistently cued adjacent dependencies is critical, then infants in the Experimental group, but not the Control group, should dishabituate to the test strings. If experience with the vocabulary elements participating in the nonadjacent dependencies is sufficient, then both Experimental and Control infants should dishabituate.

Preliminary analyses indicated that there were no differences in learning as a function of the language grammar to which infants were exposed during habituation (*t*s (30) < 1.38, *p*s < .18, and we therefore collapsed across this factor in subsequent analyses.

Paired-samples *t* tests indicated that the Experimental group listened significantly longer to the two test trials (*M* = 7.95, *SE* = .992) than to the final two habituation trials

(M = 4.59, SE = .482), t (15) = 4.571, p < .001. The Control group's listening time to

test trials (M = 6.91, SE = 1.31) and the final two habituation trials (M = 5.78, SE =

1.18) did not differ, t (15) = 1.23, p = .24. Figure 2 (in Appendix A) depicts infants'

mean listening times broken down by familiarization condition and trial type.

We next directly compared discrimination in the two groups. The Experimental

group's mean listening time to their two test trials was longer than listening to their

final two habituation trials by an average of 3.36 seconds (SE = .91), while the Control

group's mean listening to their two test trials was 1.13 seconds longer than their mean

listening to the two habituation trials (SE = .73). This difference between groups was

marginally significant, t (30) = 2.00, p = .067. Thus, the Experimental group's

discrimination for the nonadjacent dependencies was both greater than chance, and

marginally greater than the Control group's discrimination.

The difference in the two group's ability to discriminate was unrelated to

differences in the amount of their exposure to nonadjacent dependencies during

habituation. The Experimental group took an average of 8.5 trials (SE = .791) and mean

accumulated listening times of 97.02 seconds (SE = 11.16) to habituate, and the Control

group took an average of 8.3 trials (SE = .804), and mean accumulated listening times

of 94.34 seconds (SE = 12.21). Neither of these differences between familiarization

conditions were significant, ts (30) < .166, ps > .869. Additionally, the two groups did

not differ in their mean listening times to the last two habituation trials, t (30) = .934, p

= .362, indicating that their listening times at the point of habituation were comparable.

Figure 3 (in Appendix A) contains a scatter plot of habituation times for individual infants in both the Experimental and Control groups.

Our results indicate that infants are sensitive to the nonadjacent co-occurrence restrictions in the *acX bcY* language only if they are first given exposure to adjacent co-occurrences. There are two potential explanations for the Experimental group's success. First, infants' ability to detect the nonadjacent *a_X* and *b_Y* dependencies might require familiarity with those specific *aX* and *bY* pairs presented adjacently. If this were the case, infants' ability to detect the nonadjacent dependencies would represent generalization, in so far as familiarization phrases containing the adjacent dependencies differed in length and prosody from habituation phrases, and perhaps more importantly, they contained adjacent rather than nonadjacent dependencies. However, infants would be unable to generalize to novel nonadjacent combinations, and thus the benefit received from prior experience would be restricted to familiar cases. A second explanation is that Experimental infants learned something about the *aX bY* structure during familiarization (i.e. that *a*s predict bisyllabic *X*s and *b*s predict monosyllabic *Y*s), and recognized those relations in the habituation strings. This would represent even greater generalization between familiarization and habituation, and would suggest a less restricted role of prior experience. Thus, Experiment 4 tested whether infants' sensitivity to adjacent dependencies would generalize to novel nonadjacent *aX* and *bY* combinations.

Control infants were exposed to a language which was matched to the *aX bY* language in that it consisted of strings with the same prosody, and containing the

vocabulary and positional information of elements involved in the adjacent

dependencies, however it did not contain consistent co-occurrence restrictions (i.e. its

as and bs predicted the entire set of Xs and Ys). The fact that Control infants failed to

acquire sensitivity to the nonadjacent relationships they heard during habituation could

indicate that exposure to consistent $aX\ bY$ structure was necessary for such benefit.

However, there are two alternative explanations for these findings. First, our design of

the Uncued language resulted in the presence of distributional cues that as and bs

formed a single category, which we refer to as z, because in this language as and bs

predicted the entire set of Xs and Ys. If infants were sensitive to these cues, their

familiarization could have led to an abstraction involving zX and zY strings, rather than

nonoverlapping sets of aX and bY strings. By this view, habituation consisted of zcX

and zcY strings (rather than acX and bcY strings), as did test strings (rather than bcX and

acY strings), such that for Control infants habituation and test trials could both be

considered to contain strings of the same form, and therefore their failure to

dishabituate might not have reflected insensitivity to the nonadjacent relations. We

should point out that this is unlikely, as several experiments indicate that sensitivity to

the fact that a- and b- elements have nonoverlapping distributions typically relies on the

presence of cues signaling the category membership of words (Braine, 1987; Frigo and

MacDonald, 1998; Gerken et al., 2005), and the Uncued language lacked such cues. A

second alternative explanation for Control infants' lack of discrimination is that the

adjacent pairings they were exposed to during familiarization occurred nonadjacently in

both the habituation and test strings. In other words, the nonadjacent pairings they

heard at habituation had been heard adjacently during familiarization, but so had the nonadjacent pairings in the test strings. Thus, it is possible that Control infants failed to dishabituate because the nonadjacent pairings at test were familiar, rather than because they failed to notice them. To address these issues, Experiment 4 included a control language in which *a*s and *b*s predicted nonoverlapping sets of *X*s and *Y*s, and in which the nonadjacent pairings in both habituation and test strings were unfamiliar to infants.

Experiment 4

Experiment 3 demonstrated that 12-month-olds can become sensitive to nonadjacent co-occurrence relationships if they are given prior exposure to adjacent pairings. While this represents generalization of particular adjacent pairs to nonadjacent instantiations, it does not suggest that infants can generalize the more abstract co-occurrence relationships (i.e. *a*s predict bisyllabic elements and *b*s predicts monosyllabic elements) to a nonadjacent instantiation. Experiment 2 thus tested whether infants can generalize to unheard nonadjacent pairings. Specifically, both habituation and test strings contained unfamiliar nonadjacent pairings, but only those in the habituation strings were consistent with the co-occurrence restrictions in the familiarization language. A control group was exposed to an Uncued language in which there were co-occurrence restrictions, such that *a*s and *b*s predicted nonoverlapping sets of *X*s and *Y*s. However, the co-occurrence restrictions were not cued by distinctive endings and therefore infants were unlikely learn them. Like the Experimental group, they were subsequently exposed to habituation and test strings containing unfamiliar

nonadjacent pairings, with only the habituation strings containing pairings consistent with the co-occurrence restrictions present during familiarization.

Method

Participants

Participants were 32 12-month-old infants (ages ranged from 349 to 382 days) with a mean age of 367.77 days. Sixteen infants (8 female) were assigned to the Experimental familiarization condition, and 16 (8 female) were assigned to the Control familiarization condition. In both the Experimental and Control conditions, infants were randomly assigned to one of the two language grammars. As in Experiment 1, infants born prematurely, at a low-birth weight (less than 5 pounds, 8 ounces), or who had cognitive impairments or immediate family members with language delays, were not included. The data from additional infants were excluded for the following reasons; excessive fussiness (N = 24), failure to habituate (N = 5), excessive drowsiness (N = 1), failure to recover to the "recovery" trial (N = 2), parental or sibling interference (N = 8), experimenter error (N = 2), and equipment failure (N = 1).

Materials

Familiarization

The familiarization materials for the Experimental language were the same as those used for the Experiment 1 *aX bY* language, except that the number of *X*- and *Y*-elements was increased from six to eight. Increasing the number of *X*s and *Y*s allowed us to withhold a small number of phrases from familiarization, while still preserving the number of unique phrases presented in Experiment 1. This resulted in 32 unique

phrases (16 *aX* and16 *bY*) in G1, and 32 phrases (16 *aY* and 16 *bX*) in G2. However, 4

instances of each these phrase-types were withheld from familiarization in each

grammar, and thus there were 12 *aX* phrases and 12 *bY* phrases presented in G1,

(similarly, there were 12 *aY* phrases and 12 *bX* phrases in G2). Phrases from the

language were combined to form strings, each containing an *aX* phrase and a *bY* phrase

(i.e. in G1 phrases are *aX bY* or *bY aX,* and in G2 they are *aY bX* or *bX aY*). A subset of

48 of the 288 possible strings generated by combining the set of familiarization phrases

was presented during familiarization. Table 9 (in Appendix A) lists the phrases used in

the Experimental conditions.

The Uncued language consisted of 2 grammars, each containing a mixture of *aX,*

bY, aY, and *bX* strings, and containing the same vocabulary items as the Experimental

language. In G1, *a*-elements were paired with X_{1-4} and Y_{5-8}, and *b*-elements were paired

with X_{5-8} and $Y_{1-4.}$ In G2 the opposite pairings held, such that *a*-elements were paired

with X_{5-8} and Y_{1-4}, and *b*-elements were paired with X_{1-4} and Y_{5-8}. The *a*- and *b*-elements

thus predicted nonoverlapping sets, however unlike the *aX bY* language, syllable

number did not cue the sets. Because sensitivity to the fact that *a*- and *b*- elements have

nonoverlapping distributions typically relies on the presence of cues signaling the

category membership of words (Braine, 1987; Frigo and MacDonald, 1998; Gerken et

al., 2005; Gomez & Lakusta, 2004), Control infants are unlikely to learn the co-

occurrence restrictions in the Uncued language. As with the Experimental language 4

instances of each phrase-types were withheld from familiarization in each grammar. A

subset of 48 of the 288 possible strings generated by each language was presented

during familiarization. Each string contained an *aX* phrase and a *bY* phrase (i.e. in G1 phrases are *aX bY* or *bY aX*, and in G2 they are *aY bX* or *bX aY*). Table 9 lists the phrases used in each condition.

The familiarization materials were generated by the trained speaker from Experiment 1. All other aspects of both the *aX bY* and Uncued language familiarization stimuli were the same as in Experiment 3.

Habituation and Test

Habituation and test strings were created by inserting a novel *c* element into the *aX* and *bY* phrases withheld from familiarization in the Experimental and Uncued languages. Importantly, the sets of withheld strings from the Experimental and Uncued languages overlapped (see Table 9), and the subset of the strings that were grammatical for both Experimental and Control infants was used to create the habituation and test strings. It was thus possible for Experimental and Control infants to be habituated to and tested on the same set of materials containing novel nonadjacent dependencies conforming to the co-occurrence restrictions of their familiarization language (the materials are listed in Table 9 in Appendix A). There were 2 *acX* and 2 *bcY* phrases from G1 of the *acX bcY* and Uncued languages, and 2 *acY* and *bcX* phrases from G2. These phrases were combined to form 8 strings (4 *acX bcY* and 4 *bcY acX* in G1, and 4 *acY bcX* and 4 *bcX acY* in G2). We created two stimulus sets composed of different randomizations of the G1 habituation strings, and 2 sets composed of different randomizations of the G2 habituation strings. This ensured that infants were not exposed to the same order of strings on each habituation and test trial.

Procedure

Infants in both the Experimental and Control Groups were familiarized to 3 blocks (or approximately 8 minutes and 40 seconds) of strings from their familiarization language. They were subsequently habituated to strings containing nonadjacent relationships consistent with the adjacent dependencies of their familiarization language. Once they had habituated (according to the same criterion as used in Experiment 3), they were played two test trials from the other grammar (containing violations of the nonadjacent dependencies present in habituation). Half of the infants were familiarized and habituated to G1 strings, and tested with G2 strings, and the other half were familiarized and habituated to G2 strings and tested with G1 strings. All other aspects of the procedure were identical to Experiment 3.

Results and Discussion

Infants in the Experimental group were exposed to adjacent co-occurrence restrictions during familiarization, and habituated to strings containing novel nonadjacent dependencies consistent with those co-occurrence restrictions. If infants' sensitivity to the adjacent dependencies enables generalization to unfamiliar nonadjacent combinations, they should notice when they are disrupted at test (i.e. they should dishabituate). Control infants, who were unlikely to become sensitive to the co-occurrence relationships in their familiarization, should fail to detect the nonadjacent dependencies.

Preliminary analyses indicated that there were no differences in learning as a function of the grammar of the *aX bY* and Uncued languages to which infants were

exposed, t (30) = 1.04, p = .307, and this factor was therefore not included in

subsequent analyses.

Figure 4 (in Appendix A) depicts mean listening to the last two habituation trials

and the two test trials. Paired samples t tests indicated that the Experimental group's

average listening time to the two test trials (M = 8.67, SE = 1.51) was longer than

listening to the last two habituation trials (M = 5.34, SE = 1.24), t (15) = 2.61, p = .02.

Control infants' listening times to the test and habituation trials did not differ (M =

6.15, SE = .92, and M = 6.51, SE = 1.24, respectively), t (15) = .29, p = .78. An

independent samples t test indicated that the Experimental groups' difference in

listening time to habituation and test trials (M = 3.14 seconds, SE = 1.32), was

marginally greater than the difference in listening times for the Control group (M = -

.37 seconds, SE = 1.20), t (30) = 2.0, p = .054.

As in Experiment 1, the difference in the Experimental and Control groups'

sensitivity to the nonadjacent dependencies cannot be explained by differences in

exposure to them during habituation. The Experimental group accumulated an average

of 116.82 seconds (SE = 22.77) of looking over 9.13 trials (SE = 1.17), and the Control

group accumulated an average of 125.94 (SE = 22.77) seconds of listening over 8.25

trials (SE = 1.17). Neither of these differences was significant, ts (30) < .528, ps >

.601.Individual infants' habituation times can be found in Figure 5 (in Appendix A).

Experimental and Control infants were also comparable in their listening times at the

point of habituation. The Experimental group listened to the last 2 habituation trials for

an average of 5.53 seconds (*SE* = 1.24), and the Control group listened for 6.51 seconds (*SE* = 1.24), *t* (30) = .56, *p* = .581.

Discussion of Infant Experiments

In Experiments 3 and 4 we demonstrated that infants, like adults, benefit from experience with adjacent dependencies. We familiarized 12-month-old infants to a language containing adjacent dependencies, and then habituated them to a language containing nonadjacent dependencies. The familiarization and habituation phrases differed substantially in their surface characteristics; in Experiment 3 familiarization phrases, which contained adjacent dependencies (i.e. *aX* and *bY*), had a rise-fall contour over two distinct elements, and habituation phrases, which contained nonadjacent dependencies (*acX* and *bcY*) were comprised of three distinct elements with a rise-rise-fall contour. The habituation phrases, while they contained *aX* and *bY* combinations heard during familiarization, also contained an unfamiliar vocabulary element (the *c*-element). In Experiment 4, habituation further differed from familiarization phrases in that they involved unfamiliar *aX* and *bY* pairings. In both experiments, we found that infants given experience with adjacent dependencies detected the nonadjacent dependencies, as evidenced by dishabituation to strings containing violations of those relationships. The performance of infants in two Control groups, who were exposed to the vocabulary and prosody of the *aX bY* language, but not to the co-occurrence restrictions, indicated that prior experience with such surface features was not sufficient to promote sensitivity to the nonadjacent dependencies. These findings therefore suggest that prior experience can bootstrap sensitivity to complex structure, and, more

broadly, that prior experience has a significant impact on learners and the structure they are able to acquire.

Our findings also shed light on the developmental trajectory of sensitivity to nonadjacent relationships. By 18 months of age, English-learning infants demonstrate sensitivity to nonadjacent relationships between auxiliaries and inflectional endings (e.g. the sentence "The boy *is* runn*ing*" is grammatical, but "The boy *can* runn*ing*" is not) if they are separated by no more than 3 morphemes (Santelmann & Jusczyk, 1998). Gomez and Maye (2005) found that 15-month-olds, but not 12-month-olds, were able to learn similar nonadjacent relationships when exposed to simpler language materials. Such findings suggest that development of the ability to learn nonadjacent dependencies relies on maturational increases in short-term memory capacity, at least in naïve learners. However, we found that even 12-month-olds can learn nonadjacent dependencies if they are given prior experience with adjacent ones. Thus, it is not that infants younger than 15-months simply cannot detect nonadjacent relationships, but rather that they can do so *if* they can bootstrap from simpler instances of such structure. Work in many other domains of cognitive development suggest a similar role of prior experience, such that young children succeed at a task at earlier ages if they have the supporting requisite knowledge, or are given relevant scaffolding experiences (Gentner & Ratterman, 1991; Kovotsky & Gentner, 1996; Ratterman & Gentner, 1998).

Our work also speaks to infants' ability to form and extend generalizations in early language acquisition. While there is good evidence that toddlers are sensitive to abstract syntactic structure in language (e.g., Gertner, Fisher, & Eisengart, 2006; but see

Tomasello, 2000), extensive debate continues as to whether this sensitivity is innately given, or is learned through experience with statistical regularities in language input (Altmann, 2002; Altmann & Dienes, 1999; Marcus, Vijayan, Rao, & Vishton, 1999; Seidenberg & Elman, 1999). Our findings, like those of Gomez & Lakusta (2004) and Gerken et al. (2005), indicate that infants can use statistical and phonological information to acquire syntactic-like word categories and their co-occurrence dependencies. We also found that infants generalize these relations to novel instances with relatively dissimilar surface features in an artificial-language learning task, and can even generalize from these simple dependencies to complex ones that are quite difficult for naïve learners to acquire. In sum, we found that prior experience can play an important role in infants' sensitivity to difficult nonadjacent syntactic relationships, and that it can even promote such sensitivity in infants at an age at which they are typically unable to learn such relationships.

GENERAL DISCUSSION

We have demonstrated that learners' prior experience can influence the structure they are likely to acquire. This implies that the input that is required to learn difficult structure changes as learners gain experience. In particular, we found that exposure to simple structure (i.e. adjacent dependencies between word categories) facilitates learning similar but more complex structure (i.e. nonadjacent dependencies).

These data shed light on how some potentially difficult natural language structures might be acquired. One example of parallel structure in language — the co-occurrence relationship in English between determiners and nouns, and between auxiliaries and verbs — contains structure similar to that of the category induction languages used in these experiments. Specifically, dependencies between syntactic categories can be adjacent (e.g. in a "determiner-noun" string) or nonadjacent (e.g. in a "determiner-adjective-noun" string). We found that while such nonadjacent relationships are difficult for naïve learners, prior experience with similar adjacent dependencies enables such learning. Our findings are in line with some connectionist models demonstrating benefits of prior experience in learning nonadjacent syntactic dependencies. For example, Elman (1993) trained a simple recurrent network (SRN) on an artificial language with several characteristics of natural language, such that subjects and verbs agreed in number, verbs differed in their argument structure, and sentences could contain multiple relative-clause embeddings. The network succeeded in learning the language structure when it was trained first with simple sentences, and was gradually

exposed to complex strings containing more embeddings, but learning was poor without such staged input.

In related work, Conway, Ellefson, and Christiansen (2003) investigated the conditions under which staged input facilitates adults' learning of recursive structure, noting that many of the studies demonstrating an advantage of staged input used visually presented stimuli. They gave adult learners visual or auditory exposure to a language incorporating recursion. Learners given visual exposure were presented with all elements comprising a single string simultaneously, while aural presentation of string-elements was sequential. In the visual exposure condition, learners given staged input learned better than those exposed to the entirety of the language at once. Interestingly, when learners were aurally exposed, they did not benefit from staged input.

The current findings, as well and those of Elman (1993) and Conway et al. (2003), demonstrate that learning complex structure can be facilitated by prior experience with simpler structure under some circumstances. Thus, our work adds to the previous findings on the effects of staging input. However, in the languages Elman and Conway et al. Used, the same vocabulary elements were present in the simple and complex structures. Thus, they found that prior experience with particular exemplars facilitates learning when those same exemplars are involved in more complex constructions, and thus their findings show only that learners given exposure to particular adjacent combinations are better able to detect nonadjacent relationships between these same elements. Our findings with adults, however, show that prior experience facilitates

learning even when the transfer language did not involve the same vocabulary. The infant experiments do not involve far transfer, as infants' prior experience with adjacent combinations facilitated sensitivity to nonadjacent combinations involving familiar *a*, *b*, *X*, and *Y* elements. However, because the particular nonadjacent combinations had been withheld from familiarization, they indicate that the benefits to infants are not limited to heard combinations.

Interestingly, not all studies have replicated the finding that staged input benefits learning. For example, in learning some tasks, such as "exclusive or", withholding possible instantiations of a pattern can lead connectionist networks to form incorrect generalizations (Elman, 1993). In an example more relevant to the current study, Rhode and Plaut (1999) failed to replicate Elman's findings using a similar network architecture and training language. In their connectionist network (a SRN, or simple-recurrent network), information about associative relationships between elements is registered by or stored in the weights on connections between those units, and information about predictive structure is thus distributed across connections between specific items. One possibility for these findings is that initial learning of a particular set of associative or predictive relationships, is disrupted by exposure to new associations involving the same elements that are introduced in the longer, more complex strings. To translate this into terms relevant to the artificial languages used in the current study, it is possible that in connectionist networks, sensitivity to the fact that "alt" predicts "wifoo" in the string "alt wifoo" is interrupted when, at a later point in training, "alt" begins to also predict "hes" in strings like "alt hes wifoo".

Infants in our experiments, however, did benefit from exposure to adjacent relationships when exposed to nonadjacent ones involving the same elements. Our findings therefore suggest that connectionist networks with the architecture used by Elman (1993) and Rhode and Plaut (1999) may be incomplete as models of generalization. One possibility is that human learners are able to achieve a higher level, or different kind, of abstraction than connectionist networks, one that is not tied to the specific elements involved in learning.

In addition to questioning whether learning long-distance dependencies in connectionist networks is facilitated by prior exposure to local predictive relationships, Rhode and Plaut (1999) point out that in natural languages, semantic factors often constrain the kinds of material that can be embedded with different phrase types, resulting in distributional biases. Thus, the embedded material may participate in local predictive relationships or dependencies that bridge the more distant ones, and therefore the problem of learning long-distance dependencies may be ameliorated by the presence of intermediary predictive relationships. However, our data suggest that prior experience with adjacent dependencies can also play an important role in learning difficult nonadjacent relationships.

Our findings also speak to learners' ability to generalize or transfer across instances with dissimilar surface structure, a central topic in the domain of language acquisition. Recent empirical studies of infant learning indicate that infants can form generalizations relevant to natural language syntax. For example, Gomez and Lakusta (2004) demonstrated that 12-month-old infants can learn and generalize co-occurrence

relationships between word categories. Specifically, they found that infants exposed to a category induction language similar to the *aX bY* language used in Experiments 3 and 4, can discriminate grammatical and ungrammatical strings containing novel *X* and *Y* elements if they are marked by distinctive features (i.e. *X*s are bisyllabic and *Y*s are monosyllabic). Moreover, using a similar paradigm, Gerken et al. (2005) found that 18 month-old infants generalized to withheld strings even if the *X*s and *Y*s in those strings were not marked by distinctive features. The current findings extend previous work on generalization by showing that infants can generalize from adjacent to nonadjacent co-occurrence relationships, and can even do so across strings with substantial differences in their surface characteristics. Thus, these experiments add to the evidence that infants can indeed form generalizations relevant to natural language structure, and also suggest that prior experience can facilitate this process.

While we found that learners transfer to a pattern with substantially different surface characteristics but similar underlying structure, further work is needed to clarify what aspects of this similarity underlie the benefit. One possibility is the direction of predictive relationships between elements is important, as learners were exposed to two languages in which the strongest predictive relationships were prospective: the first element in the string predicted the ending on the last element with a probability of 1.0, while the ending of the final element predicted the identity of the first element with a probability of .5. If maintaining the direction of predictive relationships is critical for learners to benefit, learners transferred to a language in which the co-occurrence relationships are reversed in direction (i.e. an *Xa Yb* language) should fail to benefit, or

even show interference effects. However, we have collected preliminary data with adult

learners suggesting that learning an *aX bY* language does benefit from prior experience

with an *Xa Yb* language.

A related possibility is that similarity in the strength of predictive relationships is

important. For example, in our experiments, *a*-elements always predicted *X*-elements

marked by distinguishing features (e.g. the ending "oo"), and *b*s were always followed

by *Y*s with a different distinguishing feature (e.g. the ending "ee"). Thus, the language

structure was highly reliable, in that whenever there was an *a*-element, it was followed

by a word with the "oo" ending (and *b*s likewise always predicted the "ee" ending).

Perhaps exposure to a language that contains very high levels of predictivity facilitates

learning similar highly predictive relationships, but not more probabilistic (less reliable)

relationships. We are collecting preliminary data to address this question, and are

finding that learning a language with highly predictable co-occurrence relationships

benefits learning a new language in which such predictive relationships are less

reliable. We also found that there are limits on such benefits, as a language with low

levels of predictivity remains difficult to acquire even when learners are given

experience with a similar but more predictive structure.

Finally, in our adult experiments the features (endings on the *X* and *Y* words),

although expressed in a different surface form ("oo" and "ee" vs. "it" and "ul"), were

present in the initial and transfer language. The characteristics of the features used to

indicate category membership alone do not result in transfer (i.e. the Control group in

Experiment 3 failed to benefit from exposure a language containing words with salient

and frequently occurring endings), however what would happen if different features were used to mark the *X*- and *Y*-elements in the transfer language, such as syllable number? Preliminary data suggest that learners exposed to a an *aX bY* language in which *X*s and *Y*s are distinguished by syllable number benefit relative to naïve learners when exposed to an *aX bY* language in which *X*s and *Y*s are distinguished by distinctive endings. It is possible that if *X*s and *Y*s were distinguished along a non-acoustic dimension, such as semantic reference, learners might not transfer as readily.

Taken together, the data from this dissertation and preliminary data from studies in progress suggest that the benefits learners obtain from prior experience with an *aX bY* structure generalizes to novel languages in which there are co-occurrence restrictions on word categories, despite differences in the features used to highlight the presence of word categories, differences in the overall levels of predictivity in the languages, and differences in the direction of predictive relationships. These findings suggest that benefits from prior experience are not limited to instances in which two structures are highly similar, and indicate that prior experience can play a powerful role in facilitating subsequent learning. One concern such robust learning raises is that there are no checks on the influence that prior experience exerts over learners, however there are reasons to think this is unlikely. For example, in Experiment 2, adult learners exposed to the Uncued languages were not exposed to predictive relationships between *a*s and *b*s and the endings on *X*s and *Y*s, but they were not impaired in learning such co-occurrence restrictions in an *acX bcY* language relative to naïve learners. Thus, exposure to a language lacking detectable co-occurrence restrictions between categories

of elements does not appear to prevent learners from subsequently acquiring such structure. Moreover, it is likely that learners' prior experience with a pattern will not lead them to impose that pattern on subsequent stimuli that contain entirely different structure. Additionally, there are limits on the kinds of prior experiences that influence learners in different scenarios. For example, in a previous experiment, we found that learners exposed to an artificial language containing predictive relationships between adjacent *syllables* are no better off than naïve learners when subsequently exposed to an artificial language containing adjacent relationships between *word categories* (Lany, 2003), suggesting that transfer may only occur for structure occurring at the same linguistic level. Finally, Experiments 1 and 2 both show that learning is susceptible to interference from counterexamples: specifically, sensitivity to the *acX bcY* language diminishes over testing as learners acquire more experience with nongrammatical instances. This suggests that the incorrect generalization is likely to be overwritten given sufficient examples exhibiting a different structure, and that to the extent that learner's previous experience led them astray, it is unlikely that such learning will be maintained, let alone imposed on subsequent instances.

These findings also have implications for theories of language acquisition that incorporate a significant role for learning from language input. A growing body of studies investigating human learning mechanisms indicates considerable sensitivity to statistical information present in the environment. As previously mentioned, infants can use statistical information to learn about diverse aspects of their language. However, it is important to point out that humans are not unconstrained learners, acquiring any

regularity present in their environment. For example, Saffran (2002) demonstrated that there are modality constraints on learning predictive relationships in an artificial language. She exposed adult learners either to a phrase-structure language in which the presence of one item in a phrase predicted the presence of another item (Language P), or to a phrase-structure language in which predictive relationships were absent (Language N). Participants were presented with either an auditory or visual version of Language P or N. When presentation was auditory, learners acquired Language P better than Language N. However, learners given visual exposure acquired the predictive and non-predictive languages equally well, suggesting that learners' sensitivity to patterns and relationships varies as a function of modality.

Similarly, acquiring aurally presented nonadjacent relationships appears to be quite difficult for both infants and adults. Newport and Aslin (2004) found that adult learners readily acquire nonadjacent relationships between either vocalic or consonantal segments in fluent speech, but fail to do so when dependencies hold between nonadjacent syllables.[2] Gomez (2002) found that both infants and adults acquire nonadjacent relationships only when relationships between adjacent words were highly unpredictable (i.e. when there was high variability between the adjacent elements).

Such findings clearly demonstrate that learners come equipped with learning biases, such that certain types of structure are more salient, or more easily learned, than others. This has led to theories of language acquisition in which "innately constrained"

---

[2] Using a similar procedure to Newport and Aslin (2004), Pena, Bonatti, Nespor, and Mehler (2002) found that adults were able to acquire nonadjacent relationships between syllables in fluent speech. Many follow-up studies have taken issue with the materials, procedure, and interpretation of that study (Seidenberg, MacDonald, & Saffran, 2002; Newport & Aslin, 2004; Perruchet, Tyler, Galland, & Peereman, 2004; Onnis, Monaghan, Richmond, & Chater, 2005)

learning mechanisms play a significant role (e.g. Newport & Aslin, 2000; Saffran, 2003). The current findings provide support for this idea, in that both infant and adult learners failed to acquire co-occurrence restrictions on word categories if they held between nonadjacent categories. However, our findings also suggest that prior experience has a strong impact on such biases, as exposure to adjacent dependencies bootstrapped sensitivity to the more difficult nonadjacent ones. The input required by naïve learners for successful learning may thus differ from that required by more experienced learners. More broadly, these data support a view of language acquisition in which sensitivity to statistical information plays an important role, and in which inherent biases on learning mechanisms exist but are malleable, changing dramatically as learners' gain experience (see Elman et al., 1996; Elman, 1993).

To conclude, these findings speak to three important questions about language acquisition. First, there is growing interest in the processes underlying sensitivity to nonadjacent relationships (Gomez, 2002; Pena et al., 2002; Newport & Aslin, 2004; Santelmann & Jusczyk, 1998), as these relationships are found in natural language syntax but can be very difficult to acquire. Our findings suggest that sensitivity to such relationships can be bootstrapped from prior experience with adjacent dependencies. Second, there is still considerable controversy over learners' ability to form and generalize syntactic patterns from statistical regularities in their input. Our findings indicate that learners are capable of robust generalizations of this kind. Finally, there is currently intense interest in the nature of the learning mechanisms available to language

learners. Our findings make the important point that learning mechanisms are

powerfully influenced by learners' prior experience.

APPENDIX A

Table 1

aX bY Language Materials

Version A				Version B			
G1: aX bY				**G1: aX bY**			
G2: aY bX				**G2: aY bX**			
A	**b**	**X**	**Y**	**a**	**b**	**X**	**Y**
alt	erd	juhnoo	nusee	ush	ong	keerit	bivul
pel	vot	wifoo	lemee	dak	rud	lepit	choopul
		tamoo	sufee			feegit	habbul
		feenoo	vaymee			soolit	jerul
		zinoo	raffee			yohvit	pogul
		deechoo	durpee			zamit	vummul
G1 Grammatical-Heard Test Strings				**G1 Grammatical-Heard Test Strings**			
alt feenoo		erd vaymee		ush soolit		ong vummul	
pel zinoo		vot durpee		dak zamit		rud pogul	
G1 Grammatical-Unheard Test Strings				**G1 Grammatical-Unheard Test Strings**			
alt juhnoo		erd nusee		ush lepit		ong choopul	
pel wifoo		vot lemee		dak keerit		rud bivul	
G2 Grammatical-Heard Test Strings				**G2 Grammatical-Heard Test Strings**			
alt vaymee		erd feenoo		ush vummul		ong soolit	
pel durpee		vot zinoo		dak pogul		rud zamit	
G2 Grammatical-Unheard Test Strings				**G2 Grammatical-Unheard Test Strings**			
alt nusee		erd juhnoo		ush choopul		ong lepit	
pel lemee		vot wifoo		dak bivul		rud keerit	

Table 2

acX bcY Language Materials

Version A				
G1: acX bcY				
G2: acY bcX				
a	**b**	**c**	**X**	**Y**
alt	erd	hes	juhnoo	nusee
pel	vot	kaf	wifoo	lemee
		sij	tamoo	sufee
			feenoo	vaymee
			zinoo	raffee
			deechoo	durpee

G1 Grammatical-Heard Test Strings			
alt hes feenoo	pel hes zinoo	erd hes vamee	vot hes rafee
alt kaf wifoo	pel kaf deechoo	erd kaf rafee	vot kaf durpee
alt sij tamoo	pel sij juhnoo	erd sij durpee	vot sij nusee

G1 Grammatical Unheard Test Strings			
alt hes juhnoo	pel hes wifoo	erd hes nusee	vot hes lemee
alt kaf wifoo	pel kaf tamoo	erd kaf lemee	vot kaf sufee
alt sij tamoo	pel sij feenoo	erd sij sufee	vot sij vamey

G2 Grammatical-Heard Test Strings			
alt hes vamee	pel hes rafee	erd hes feenoo	vot hes zinoo
alt kaf rafee	pel kaf durpee	erd kaf zinoo	vot kaf deechoo
alt sij durpee	pel sij lemee	erd sij zinoo	vot sij juhnoo

G2 Grammatical-Unheard Test Strings			
alt hes nusee	pel hes lemee	erd hes juhnoo	vot hes wifoo
alt kaf lemee	pel kaf sufee	erd kaf wifoo	vot kaf tamoo
alt sij sufee	pel sij vaymee	erd sij tamoo	vot sij feenoo

Version B

G1: acX bcY

G2: acY bcX

a	b	c	X	Y
ush	ong	tash	keerit	bivul
dak	rud	fis	lepit	choopul
		nep	feegit	habbul
			soolit	jerul
			yohvit	pogul
			zamit	vummul

G1 Grammatical-Heard Test Strings

ush tash soolit	dak tash yohvit	ong tash jerul	rud tash pogul
ush fis yohvit	dak fis zamit	ong fis pogul	rud fis vummul
ush nep zamit	dak nep keerit	ong nep vummul	rud nep bivul

G1 Grammatical Unheard Test Strings

ush tash keerit	dak tash lepit	ong tash bivul	rud tash choopul
ush fis lepit	dak fis feegit	ong fis choopul	rud fis habbul
ush nep feegit	dak nep soolit	ong nep habbul	rud nep jerul

G2 Grammatical-Heard Test Strings

ush tash jerul	dak tash pogul	ong tash soolit	rud tash yohvit
ush fis pogul	dak fis vummul	ong fis yohvit	rud fis zamit
ush nep vummul	dak nep bivul	ong nep zamit	rud nep kirit

G2 Grammatical-Unheard Test Strings

ush tash bivul	dak tash choopul	ong tash keerit	rud tash lepit
ush fis choopul	dak fis habbul	ong fis lepit	rud fis feegit
ush nep habbul	dak nep jerul	ong nep feegit	rud nep soolit

Table 3

Experiment 1 Mean Endorsement Rates with Standard Error of the Mean in

Parentheses

	Block 1				Block 2			
	GH	GUH	NGH	NGUH	GH	GUH	NGH	NGUH
Control	.70 (.039)	.69 (.037)	.72 (.037)	.68 (.040)	.58 (.037)	.61 (.033)	.59 (.031)	.51 (.033)
Transfer	.71 (.031)	.68 (.035)	.62 (.040)	.55 (.042)	.63 (.040)	.60 (.040)	.58 (.040)	.53 (.037)

Note: GH refers to grammatical-heard test strings, and NGH to their matched nongrammatical

counterparts. Similarly, GUH refers to grammatical-unheard test strings, and NGH to their

nongrammatical counterparts.

Table 4

Experiment 1 Mean Discrimination for GH and GUH strings in terms of Difference

Scores and d', with Standard Error of the Mean in Parentheses

	Block 1				Block 2			
	GH		GUH		GH		GUH	
	GH-NGH	*d'*	GUH-NGUH	*d'*	GH-NGH	*d'*	GUH-NGUH	*d'*
Control	-.02 (.051)	-.36 (.312)	.01 (.049)	.03 (.342)	-.02 (.045)	.03 (.312)	.10 (.037)	.34 (.238)
Transfer	.09 (.049)	.32 (.312)	.13 (.051)	.71 (.321)	.04 (.060)	.34 (.275)	.08 (.052)	.65 (.238)

Note: GH refers to grammatical-heard test strings, and NGH to their matched nongrammatical counterparts. Similarly, GUH refers to grammatical-unheard test strings, and NGUH to their nongrammatical counterparts.

Table 5

Uncorrelated Language Materials

Version A				Version B			
G1: aX$_{1-3}$ bY$_{4-6}$				**G1: aX$_{1-3}$ bY$_{4-6}$**			
G2: aY$_{1-3}$ bX$_{4-6}$				**G2: aY$_{1-3}$ bX$_{4-6}$**			
a	**b**	**X$_{1-3}$, Y$_{4-6}$**	**X$_{4-6}$, Y$_{1-3}$**	**a**	**b**	**X$_{1-3}$, Y$_{4-6}$**	**X$_{4-6}$, Y$_{1-3}$**
alt	erd	juhnoo	nusee	ush	ong	keerit	feegit
pel	vot	wifoo	lemee	dak	rud	lepit	soolit
		feenoo	durpee			zamit	yohvit
		sufee	tamoo			bivul	habbul
		vaymee	zinoo			choopul	jerul
		raffee	deechoo			pogul	vummul
G1 Grammatical-Heard Test Strings				**G1 Grammatical-Heard Test Strings**			
alt feenoo		erd vamey		ush pogul		ong vummul	
pel zinoo		vot durpee		dak lepit		rud soolit	
G1 Grammatical-Unheard Test Strings				**G1 Grammatical-Unheard Test Strings**			
alt juhnoo		erd nusee		ush kirit		ong feegit	
pel wifoo		vot lemee		dak bivul		rud habbul	
G2 Grammatical-Heard Test Strings				**G2 Grammatical-Heard Test Strings**			
alt vamey		erd feenoo		ush vummul		ong pogul	
pel durpee		vot zinoo		dak soolit		rud lepit	
G2 Grammatical-Unheard Test Strings				**G2 Grammatical-Unheard Test Strings**			
alt nusee		erd juhnoo		ush feegit		Ong kirit	
pel lemee		vot wifoo		dak habbul		rud bivul	

Table 6

Experiment 2 Mean Endorsement Rates with Standard Error of the Mean in Parentheses

	Block 1				Block 2			
Day 1	GH	GUH	NGH	NGUH	GH	GUH	NGH	NGUH
Control	.78 (.044)	.68 (.038)	.68 (.048)	.64 (.050)	.67 (.040)	.61 (.043)	.65 (.039)	.57 (.045)
Transfer	.86 (.031)	.76 (.044)	.64 (.057)	.51 (.064)	.78 (.035)	.62 (.046)	.58 (.052)	.53 (.049)
Day 2								
Control	.64 (.030)	.56 (.042)	.63 (.038)	.60 (.036)	.63 (.042)	.62 (.036)	.62 (.037)	.55 (.043)
Transfer	.73 (.031)	.68 (.043)	.57 (.048)	.53 (.046)	.62 (.044)	.60 (.035)	.53 (.045)	.53 (.046)

Note: GH refers to grammatical-heard test strings, and NGH to their matched nongrammatical counterparts. Similarly, GUH refers to grammatical-unheard test strings, and NGH to their nongrammatical counterparts.

Table 7

Experiment 2 Mean Discrimination for GH and GUH strings in terms of Difference Scores and d', with

Standard Error of the Mean in Parentheses

	Block 1				Block 2			
	GH		GUH		GH		GUH	
	GH-NGH	d'	GUH-NGUH	d'	GH-NGH	d'	GUH-NGUH	d'
Day 1								
Control	.11 (.060)	.69 (.396)	.04 (.064)	.03 (.383)	.02 (.054)	-.03 (.361)	.04 (.070)	.27 (.373)
Transfer	.22 (.062)	1.45 (.481)	.25 (.081)	1.86 (.545)	.20 (.065)	1.25 (.459)	.08 (.076)	.46 (.490)
Day 2								
Control	.01 (.047)	-.13 (.306)	-.04 (.055)	-.21 (.344)	.01 (.051)	.18 (.360)	.07 (.045)	.51 (.307)
Transfer	.16 (.064)	.91 (.306)	.15 (.066)	.89 (.344)	.09 (.075)	.73 (.306)	.07 (.066)	.41 (.307)

Note: GH refers to grammatical-heard test strings, and NGH to their matched

nongrammatical counterparts. Similarly, GUH refers to grammatical-unheard test

strings, and NGH to their nongrammatical counterparts.

Table 8: Experiment 3 Language Materials

Familiarization Materials			
G1			
a₁X	**a₂X**	**b₁Y**	**b₂Y**
ong coomo	erd coomo	alt deech	ush deech
ong fengle	erd fengle	alt ghope	ush ghope
ong kicey	erd kicey	alt jic	ush jic
ong loga	erd loga	alt skige	ush skige
ong paylig	erd paylig	alt vabe	ush vabe
ong wazil	erd wazil	alt tam	ush tam
G2			
a₁Y	**a₂Y**	**b₁X**	**b₂X**
ong deech	erd deech	alt coomo	ush coomo
ong ghope	erd ghope	alt fengle	ush fengle
ong jic	erd jic	alt kicey	ush kicey
ong skige	erd skige	alt loga	ush loga
ong vabe	erd vabe	alt paylig	ush paylig
ong tam	erd tam	alt wazil	ush wazil

Habituation and Test Materials			
G1			
a₁X	**a₂X**	**b₁Y**	**b₂Y**
ong hes coomo	erd hes coomo	alt hes deech	ush hes deech
ong hes fengle	erd hes fengle	alt hes ghope	ush hes ghope
ong hes kicey	erd hes kicey	alt hes jic	ush hes jic
ong hes loga	erd hes loga	alt hes skige	ush hes skige
ong hes paylig	erd hes paylig	alt hes vabe	ush hes vabe
ong hes wazil	erd hes wazil	alt hes tam	ush hes tam
G2			
a₁Y	**a₂Y**	**b₁X**	**b₂X**
ong hes deech	erd hes deech	alt hes coomo	ush hes coomo
ong hes ghope	erd hes ghope	alt hes fengle	ush hes fengle
ong hes jic	erd hes jic	alt hes kicey	ush hes kicey
ong hes skige	erd hes skige	alt hes loga	ush hes loga
ong hes vabe	erd hes vabe	alt hes paylig	ush hes paylig
ong hes tam	erd hes tam	alt hes wazil	ush hes wazil

Note: Infants in the Experimental group were familiarized to G1 or G2 strings, and then heard habitation strings from the same grammar. At test they heard the habituation strings from the other grammar. Infants in the Control group were familiarized to both G1 and G2

strings, and then heard exclusively G1 or G2 habituation strings. At test they heard

habituation strings from the other grammar.

Table 9: Experiment 4 Language Materials

Familiarization Materials							
***aX bY* Language**				**Random Language**			
G1				**G1**			
				a_1X_{1-4}	a_2X_{1-4}	b_1Y_{1-4}	b_1Y_{1-4}
a_1X_{1-8}	a_2X_{1-8}	b_1Y_{1-8}	b_2Y_{1-8}	a_1Y_{5-8}	a_2Y_{5-8}	b_1X_{5-8}	b_1X_{5-8}
ong coomo	erd coomo	*alt deech*	ush deech	*ong coomo*	erd coomo	*alt deech*	ush deech
ong fengle	*erd fengle*	alt ghope	*ush ghope*	ong fengle	*erd fengle*	alt ghope	*ush ghope*
ong kicey	erd kicey	alt jic	ush jic	ong kicey	erd kicey	alt jic	ush jic
ong loga	erd loga	alt skige	ush skige	ong loga	erd loga	alt skige	ush skige
ong paylig	erd paylig	*alt vabe*	ush vabe	*ong vabe*	erd vabe	*alt paylig*	ush paylig
ong wazil	*erd wazil*	alt tam	*ush tam*	ong tam	*ong tam*	alt wazil	*ush wazil*
ong bevit	erd begit	alt vot	ush vot	ong vot	erd vot	alt bevit	ush bevit
ong meeper	erd meeper	alt rud	ush rud	ong rud	erd rud	alt meeper	ush meeper
G2				**G2**			
				a_1Y_{1-4}	a_2Y_{1-4}	b_1X_{1-4}	b_1X_{1-4}
a_1Y_{1-8}	a_2Y_{1-8}	b_1X_{1-8}	b_2X_{1-8}	a_1X_{5-8}	a_2X_{5-8}	b_1Y_{5-8}	b_1Y_{5-8}
ong deech	erd deech	*alt coomo*	ush coomo	*ong ghope*	erd deech	*alt coomo*	ush coomo
ong ghope	*erd ghope*	alt fengle	*ush fengle*	ong ghope	*erd ghope*	alt fengle	*ush fengle*
ong jic	erd jic	alt kicey	ush kicey	ong jic	erd jic	alt kicey	ush kicey
ong skige	erd skige	alt loga	ush loga	ong skige	erd skige	alt loga	ush loga
ong vabe	erd vabe	*alt paylig*	ush paylig	*ong paylig*	erd paylig	*alt vabe*	ush vabe
ong tam	*erd tam*	alt wazil	*ush wazil*	ong wazil	*erd wazil*	alt tam	*ush tam*
ong vot	erd vot	alt bevit	ush bevit	ong bevit	erd begit	alt vot	ush vot
ong rud	erd rud	alt meeper	ush meeper	ong meeper	erd meeper	alt rud	ush rud
Habituation and Test Materials							
G1				**G2**			
ong hes coomo		alt hes deech		ong hes deech		alt hes coomo	
erd hes fengle		erd hes ghope		erd hes ghope		ush hes fengle	

Note: In both the Experimental and Control groups, infants familiarized to G1 or G2
strings from their respective familiarization language, and then heard habitation
strings from the same grammar. At test they heard the habituation strings from the
other grammar. Strings withheld from familiarization are in italics.

Figure 1. A typical *aX bY* pattern. Learners are exposed to a subset of the

grammatical pairings of markers and content-words. Learners are tested to see if they

will generalize correctly to the withheld pairings.

	X_1	X_2	X_3	X_4	X_5	X_6
a_1	a_1X_1	a_1X_2	a_1X_3	a_1X_4	a_1X_5	a_1X_6
a_2	a_2X_1	a_2X_2	a_2X_3	withheld	a_2X_5	a_2X_6

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
b_1	$b_1 Y_1$	$b_1 Y_2$	$b_1 Y_3$	withheld	$B_1 Y_5$	$b_1 Y_6$
b_2	$b_2 Y_1$	$b_2 Y_2$	$b_2 Y_3$	$b_2 Y_4$	$B_2 Y_5$	$b_2 Y_6$

Figure 2. Experiment 3 Habituation and Test. Infants' mean listening times to the last two habituation trials and the two test trials in the Experimental and Control Groups.
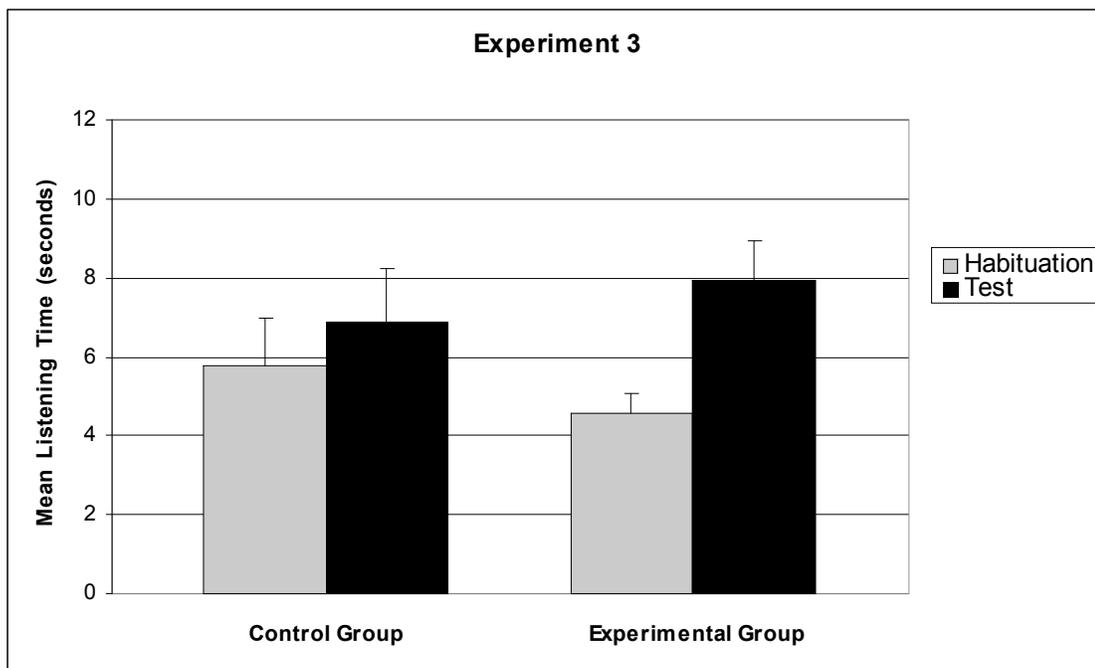
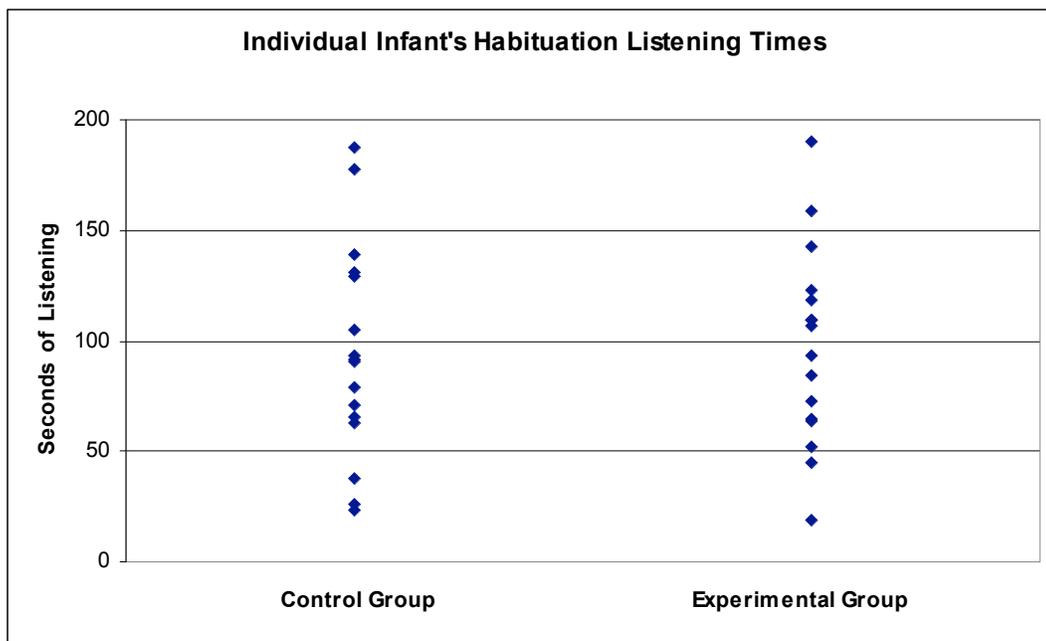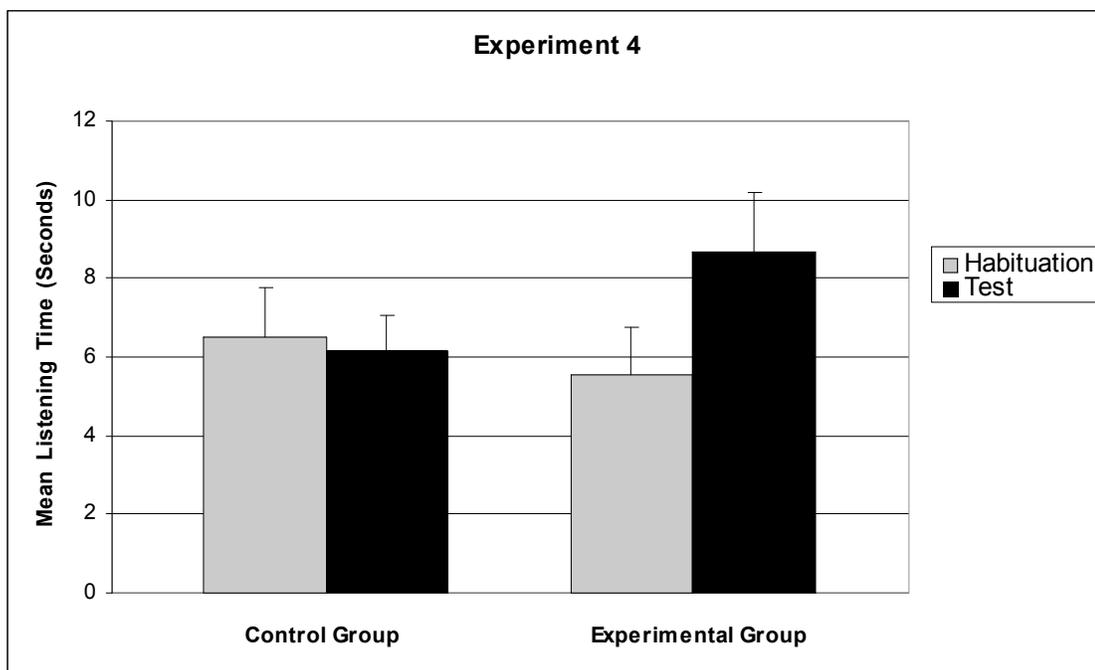Figure 3. Experiment 3 listening times during habituation for Control and Experimental infants.

Figure 4. Experiment 4 Habituation and Test. Infants' mean listening times to the last two habituation trials and the two test trials in the Experimental and Control Groups.

Figure 5. Experiment 4 listening times during habituation for Control and Experimental infants.
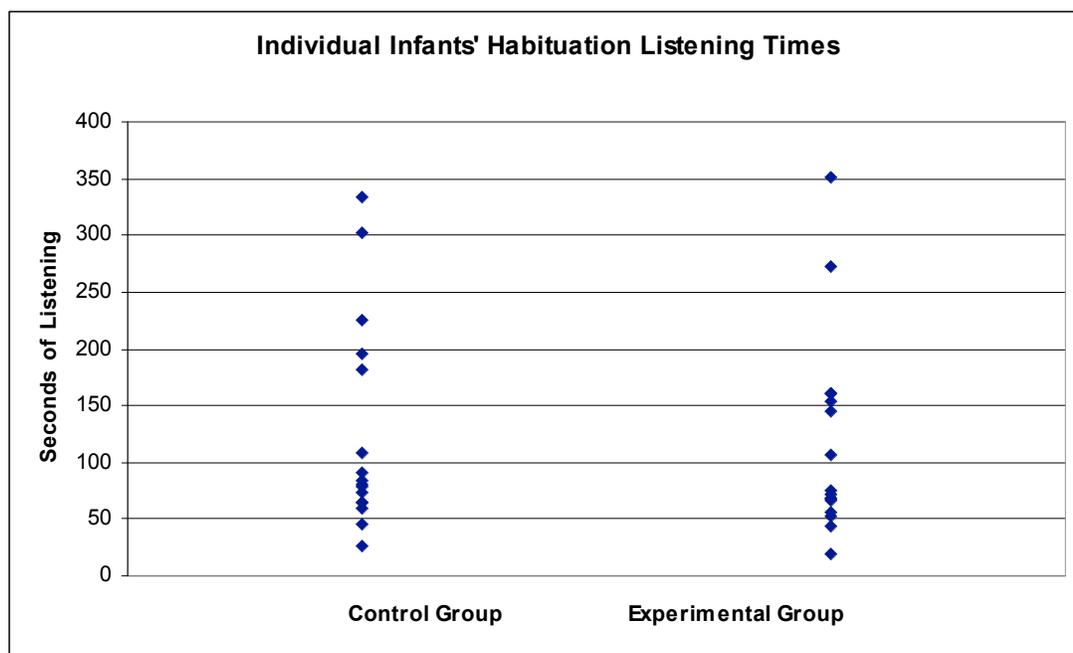
REFERENCES

Altmann, G.T.M., Dienes, Z., Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 899-912.

Altmann, G.T.M. (2002). Learning and development in neural networks – the importance of prior experience. *Cognition, 85,* B43-B50.

Altmann, G.T.M., & Dienes, Z. (1999). *Science, 284,* 875.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321-324.

Baker, M.C. (2001). *The atoms of language.* New York: Basic Books.

Bernstein-Ratner, N. (1984). Patterns of vowel modification in motherese. *Journal of Child Language, 11,* 557-578.

Braine, M.D.S. (1987). What is learned in acquiring word classes – A step toward an acquisition theory. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition.* Hillsdale, NJ: Erlbaum.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, the Netherlands: Foris Publications.

Cohen, L.B., Atkinson, D.J., & Chaput, H.H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin: University of Texas.

Conway, C.C., Ellefson, M.R., & Christiansen, M.H. (2003). When less is less and when less is more: Starting small with staged input. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71-99.

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition, 82,* 259-278.

Frigo L., & MacDonald, J. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39,* 448-457.

Gerken, L.A. (2004). Nine-month-olds extract structural principles required for natural language. *Cognition, 93*, B89-B96.

Gerken, LA., Wilson, R., & Lewis, W. (2005). Seventeen-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249-268.

Goldowsky, B.N., &, Newport, E.J. (1993). Modeling the effects of processing limitations on the acquisition of morphology: The less is more hypothesis. In E. Clark (Ed.), *The Proceedings of the 24th Annual Child Language Research Forum. Center for the Study of Language and Information*, Stanford, CA, pp. 124-138.

Gomez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7,* 183-206.

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 Science, 4*, 178-186.

Gómez, R.L., & LaKusta, L. (2004). A first step in form-based category abstraction in 12-month-old infants. *Developmental Science*, 7, 567-580.

Gómez, R.L., (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology, 33,* 154-207.

Gómez, R.L., Gerken, L.A., & Schvaneveldt, R. (2000). The basis of transfer in artificial grammar learning. *Memory and Cognition, 28,* 253-263.

Hudson Kam, C.L., & Newport, E.L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151-195.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5,* 65-95.

Lany, J. (2003). The Role of Prior Experience in Language Acquisition. Master's Thesis, The University of Arizona.

Lany, J., Gomez, R.L., and Gerken, L.A. (in press). The role of prior experience in language acquisition. *Cognitive Science.*

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Marcus, G.F.; Vijayan, S., Bandi Rao, S., & Vishton, P.M., (1999). Rule learning by seven-month-old infants. *Science, 283,* 77-80.

Maye, J., Werker, J. F., & Gerken, L.A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* B101-B111.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

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

Newport, E.L., Gleitman, H., & Gleitman, L.R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. Snow & C. Ferguson (Eds.) *Talking to Children: Language Input and Acquisition*. Cambridge University Press, Cambridge, pp. 109-149.

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

Perruchet, P., Peereman, R., & Tyler, M.D. (2006). Do we need algebraic-like computations? A reply to Bonatti, Pena, Nespor, & Mehler (2006). *Journal of Experimental Psychology: General, 135,* 322-326.

Perruchet, P., Tyler, M.D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133,* 573-583.

Pine, J.M. (1994). The language of primary caregivers. In C. Gallaway & C.J. Richards (Eds.) *Input and Interaction in Language Acquisition*. Cambridge University Press, Cambridge, pp.38-55.

Pine, J.M. (1994). The language of primary caregivers. In C. Gallaway & C.J. Richards (Eds.) *Input and Interaction in Language Acquisition*. Cambridge University Press, Cambridge, pp.38-55.

Reber, A.S. (1969). Transfer of synthetic structure in synthetic languages. *Journal of Experimental Psychology: General, 120,* 112-116.

Rohde, D.L.T. & Plaut, D.C (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition, 72*, 67-109.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4,* 273-274.

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

Saffran, J.R., & Thiessen, E.D. (2003). Patten induction by infant language learners. *Developmental Psychology, 39,* 483-494.

Saffran, J.R., (2001), Words in a sea of sound: The output of statistical learning. *Cognition, 81,* 149-169.

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

Saffran, J.R., Newport, E.L,. & Aslin, R.N., (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

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

Seidenberg, M.S., & Elman, J.L. (1999). Do infants learn grammar with algebra or statistics? *Science, 284,* 433.

Seidenberg, M.S., Macdonald, M.C., & Saffran, J.R. (2002). Does grammar start where statistics stop? *Science, 298,* 553-554.

Thiessen, E.D., & Saffran, J.R. (2003). When cues collide: use of stress and statistical cues to word boundaries. *Developmental Psychology, 39*, 706-716.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74,* 209-253.

Tunney, R. & Altmann, G. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology, 27,* 614-639.

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

Wilson, R. (2002). *Syntactic category learning in a second language*. Unpublished Ph.D. Dissertation, University of Arizona, Tucson, AZ.