Published in final edited form as:
*Psychol Sci*. 2008 December ; 19(12): 1247–1252. doi:10.1111/j.1467-9280.2008.02233.x.

# Twelve-Month-Old Infants Benefit From Prior Experience in Statistical Learning

**Jill Lany**[1] and **Rebecca L. Gómez**[2]

[1] University of Wisconsin–Madison

[2] University of Arizona

## Abstract

A decade of research suggests that infants readily detect patterns in their environment, but it is unclear how such learning changes with experience. We tested how prior experience influences sensitivity to statistical regularities in an artificial language. Although 12-month-old infants learn adjacent relationships between word categories, they do not track nonadjacent relationships until 15 months. We asked whether 12-month-old infants could generalize experience with adjacent dependencies to nonadjacent ones. Infants were familiarized to an artificial language either containing or lacking adjacent dependencies between word categories and were subsequently habituated to novel nonadjacent dependencies. Prior experience with adjacent dependencies resulted in enhanced learning of the nonadjacent dependencies. Female infants showed better discrimination than males did, which is consistent with earlier reported sex differences in verbal memory capacity. The findings suggest that prior experience can bootstrap infants' learning of difficult language structure and that learning mechanisms are powerfully affected by experience.

---

Statistical information, such as the frequency of different events and the likelihood of their co-occurrence with other events, provides us with critical knowledge about our environment. For example, facial features are likely to co-occur (e.g., mouths co-occur with two eyes and a nose). Likewise, the occurrence of the word *a* or *the* suggests that a noun like *cat* or *duck* is likely to follow. Sensitivity to statistical information emerges early in development, with infants detecting regularities in visual input by 2 months of age (Kirkham, Slemmer, & Johnson, 2002) and in aurally presented stimuli by 7 months (Saffran, Aslin, & Newport, 1996; Thiessen & Saffran, 2003). Research over the last decade suggests that statistical learning plays a critical role in language acquisition by supporting segmentation of words from fluent speech (Saffran et al., 1996), acquisition of word-order patterns (Gómez & Gerken, 1999), and syntactic categories and their co-occurrence privileges (Gerken, Wilson, & Lewis, 2005; Gómez & LaKusta, 2004).

One important question researchers would like to answer is how statistical learning is affected by experience, and they have begun to address this question in the domains of word segmentation and vocabulary development. For example, experience with phonological properties helps infants segment words from fluent speech (Saffran & Thiessen, 2003; Thiessen & Saffran, 2007), and successful word segmentation in turn facilitates learning co-occurrence relationships between words (Saffran & Wilson, 2003). Likewise, experience with associations between individual objects and labels can attune infants to higher order correspondences between words and their referents (e.g., object labels tend to refer to things

Address correspondence to Jill Lany, Waisman Center, University of Wisconsin–Madison, 1500 Highland Ave., Madison, WI 53705, lany@wisc.edu.

with similar shapes) that facilitate the rapid formation of new object-label associations (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002).

In our study, we asked how prior experience influences the acquisition of complex syntactic patterns. For example, sensitivity to nonadjacent relationships between words is necessary for learning the hierarchical structure of languages (e.g., the determiner–noun relationship is separated by an adjective in *the yellow ducky*; Chomsky, 1957). Although infants and adults readily detect adjacent relationships, nonadjacent relationships are extremely difficult to acquire (Gómez, 2002; Newport & Aslin, 2004). Lany, Gómez, and Gerken (2007) found that exposing adults to adjacent dependencies greatly facilitates their ability to learn nonadjacent dependencies, suggesting that scaffolding from simpler forms is one way these relationships are acquired (see also Elman, 1993; Newport, 1990; although see Rhode & Plaut, 1999).

A critical question is whether infants are similarly influenced by prior experience. By 12 months, infants can learn dependencies between adjacent word categories (Gómez & LaKusta, 2004), which may in turn attune them to nonadjacent dependencies. However, this requires generalizing over changes in the surface features of the strings containing such dependencies, and it is unclear whether infants are capable of such generalization. For example, there is ongoing debate over whether children have item-specific (Tomasello, 2000) or category-level representations of syntactic structures early on (Fisher, 2002; Gertner, Fisher, & Eisengart, 2006).

We therefore tested whether experience with adjacent dependencies can bootstrap infants' ability to track nonadjacent ones. Although 15-month-old infants successfully track non-adjacent relationships in an artificial language, 12-month-old infants fail to do so (Gómez & Maye, 2005), but what happens if 12-month-old infants are exposed to adjacent dependencies before they are tested on novel nonadjacent relationships?

## METHOD

### Participants

The participants in this experiment were 32 infants with a mean age of 367.77 days (range = 349–382 days). Sixteen infants (50% female) were assigned to both the experimental and control conditions, with language version (V1, V2) counterbalanced across conditions. We excluded infants of fewer than 37 weeks gestation; those who weighed less than 5 lb 8 oz at birth; those who had immediate family members with language disorders or delays; and those who were being treated for ear infections. Data from additional infants were excluded for fussiness (*n* = 24), failure to habituate (*n* = 5), drowsiness (*n* = 1), failure to recover to the "recovery" trial (*n* = 2), parental interference (*n* = 8), experimenter error (*n* = 2), and equipment failure (*n* = 1). Exclusion rates were comparable across condition and sex: *n* = 19 (9 female, 10 male) in the experimental condition, and *n* = 24 (11 female, 13 male) in the control condition.

### Materials

**Familiarization Stimuli—**Infants in the experimental group were exposed to an artificial language consisting of the word categories *a*, *b*, *X*, and *Y*, with restrictions on how words were combined into two-word phrases (similar to the co-occurrence relationships between determiners and nouns in English). The materials were adapted from Gómez and LaKusta (2004), as 12-month-old infants successfully learn its co-occurrence relationships (see Fig. 1). There were two monosyllabic *a*s and *b*s, eight disyllabic *X*s, and eight monosyllabic *Y*s. There were two versions of the language, with *aX* and *bY* pairings in V1 and *aY* and *bX* pairings in V2. The language contained distributional information that cued word categories, with *a* and *b* elements predicting nonoverlapping sets. The phonological feature distinguishing *X*s from *Y*s (syllable number) is referred to as a *correlated cue* because it is redundant with the distributional information. Natural languages often incorporate correlated cues (e.g., Monaghan, Chater, & Christiansen, 2005), and infants fail to learn *aX bY* patterns without these cues (Gerken et al., 2005; Gómez & LaKusta, 2004).

In the experimental materials, V1 generated 16 *aX* phrases and 16 *bY* phrases. Four instances of each phrase type were withheld from familiarization to test generalization to nonadjacent combinations. Phrases were combined to form strings containing one *aX* phrase and one *bY* phrase (V2 generated *aY* and *bX* phrases). A subset of 48 of the 288 possible strings generated by each language was presented during familiarization.

The control language materials contained the same vocabulary elements as the experimental language, but in V1, *a* elements were paired with $X_{1–4}$ and $Y_{5–8}$, and *b* elements were paired with $X_{5–8}$ and $Y_{1–4}$, and in V2, *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 experimental condition, the sets were not uniquely cued by syllable number. We withheld two of each of the *aX*, *aY*, *bX*, and *bY* phrases from each version to test generalization. The remaining phrases were combined to form strings containing either an *aX* and *bY* phrase or an *aY* and *bX* phrase.

The control and experimental languages were thus equated on several critical dimensions: the individual *a*, *b*, *X*, and *Y* elements were presented with equal frequency; *a*s and *b*s always occurred in string initial position, with *X*s and *Y*s always string final; and phrases consisted of two elements with a rise–fall prosodic contour. However, because correlated cues are needed to learn that *a*s and *b*s predict different sets, control infants should fail to detect the adjacent co-occurrence restrictions.

**Habituation and Test Stimuli—**Habituation and test strings were created by inserting a novel *c* element (*hes*) into withheld phrases that were grammatical for both experimental and control infants (see Fig. 1). Thus, the *aX* and *bY* dependencies were nonadjacent in this phase. There were two *acX* and two *bcY*phrases from V1, and two *acY*and two *bcX* phrases from V2. These phrases were combined to form four *acX bcY*phrases and four *bcYacX* phrases in V1 and four *acY bcX* and four *bcX acY* in V2. We created two randomized stimulus sets of the V1 and V2 habituation strings, each consisting of two randomized blocks.

Language materials were spoken by a female speaker and digitized for editing. The familiarization phrases had rising intonation over the first word and falling intonation over the second. Words within phrases were separated by 0.03 s of silence. The same word tokens were used to create the experimental and control languages, thus matching their prosodic and phonological characteristics. Habituation and test phrases had rising intonation over the first two words and falling intonation over the third. Familiarization and habituation strings consisted of two phrases separated by 0.4 s of silence. Strings were separated by 1.1 s of silence.

## Procedure

**Familiarization Phase—**The purpose of the familiarization phase was to expose infants in the experimental group to adjacent dependencies between word categories and to expose infants in the control group to a language lacking such dependencies. Familiarization took place in a playroom. Infants were familiarized to three blocks (8 min, 40 s) of their training language, each of which contained the 48 familiarization strings in a different random order.

Thus, each string was presented three times, once per block. Infants could play with toys in the playroom and with their parents, who were instructed not to talk, but to otherwise interact naturally with their infants.

**Habituation Phase—**After familiarization, infants were taken to a booth and seated on their parent's lap. An experimenter outside the booth monitored and recorded the infants' looking behavior through a TV monitor, controlling the habituation procedure using Habit X software (Cohen, Atkinson, & Chaput, 2004). Habituation trials began with a bull's-eye projected onto a screen in front of the infant. Once the infant fixated the bull's-eye, an animation of a bouncing baby appeared, and habituation strings containing nonadjacent dependencies began to play from a speaker centered above the screen. The image and auditory stimulus persisted until infants looked away from the screen for 2 consecutive seconds or until the trial had lasted 60 s. At this point, a new trial began. This process repeated until an infant's summed listening time across three trials fell to 50% of their listening time for the first three trials. There were eight unique habituation strings, each lasting approximately 3.6 s. A 60-s habituation trial thus consisted of 13 strings, with each string occurring twice, at most. On each trial, infants heard one of the two randomized sets of habituation strings. Infants familiarized to V1 of the experimental or control languages heard V1 habituation strings, and those familiarized to V2 languages heard V2 habituation strings.

**Test Phase—**After habituation, infants were exposed to two test trials containing violations of the nonadjacent dependencies in the habituation strings. Infants habituated to V1 *acX bcY* strings heard V2 *acY bcX* strings during the test, and those habituated to V2 strings heard V1 strings. Habituation and test strings contained the same adjacent *ac*, *bc*, *cX*, and *cY* transitions and thus could be distinguished only by their nonadjacent dependencies.

Finally, infants were exposed to a recovery trial that differed from test trials only in that the auditory stimuli consisted of the novel syllables *boo* and *baa*. Infants failing to listen longer to this trial than to their last two habituation trials had presumably stopped paying attention to the auditory stimuli, and their data were excluded. Parents listened to masking music over headphones throughout habituation and test phases and were instructed to remain unresponsive to the visual stimuli presented on the screen.

## RESULTS

Discrimination was measured as a difference between the mean listening times for the two test trials and the mean listening times for the final two habituation trials, with a significant increase in listening time for the test trials indicating discrimination of the nonadjacent dependencies. Preliminary analyses revealed no differences in learning as a function of language version, $t(30) = 1.04$, $p = .307$, and this factor was not included in subsequent analyses.

An analysis of variance with familiarization condition and sex as between-participant factors revealed that the experimental group's discrimination exceeded the control group's ($M$ = 3.14 s, $SE$ = 1.13 and $M$ = −0.37 s, $SE$ = 1.13 respectively), $F(1, 28) = 4.79$, $p = .037$, $\eta = .146$ (see Fig. 2). We also found that females showed greater discrimination than males ($M$ = 3.63 s, $SE$ = 1.19 vs. $M$ = −0.85 s, $SE$ = 1.18), $F(1, 28) = 7.83$, $p = .009$, $\eta = .218$. There was no interaction between condition and sex. Two-tailed *t* tests indicated that females in the experimental group showed significant discrimination of the nonadjacent dependencies ($M$ = 5.5 s, $SE$ = 2.01), $t(7) = 2.73$, $p = .029$, whereas males did not ($M$ = 0.78 s, $SE$ = 0.75), $t(7) = 1.04$, $p = .33$. Likewise, the difference in discrimination between experimental females and males was significant, $t(14) = 2.2$, $p = .045$. Control infants failed to discriminate regardless of sex ($M = 1.75$ s, $SE = 1.0$ and $M = -2.48$ s, $SE = 2.15$, for females and males respectively), $ts(7) \leq 1.75$, $ps \geq .12$. Testing 4 additional males in the experimental condition to increase power did not change the pattern of findings (see Fig. 2).

Finally, the difference in discrimination between groups was unrelated to differences in the amount of exposure to nonadjacent dependencies during habituation (see Table 1).

## DISCUSSION

We tested whether prior experience facilitates learning nonadjacent relationships, which are prevalent in natural language but difficult to acquire (Gómez, 2002; Newport & Aslin, 2004). Although infants typically fail to track nonadjacent dependencies at 12 months, they succeed at 15 to 18 months, a development that has been hypothesized to rely on maturational increases in short-term memory capacity (Gómez & Maye, 2005; Santelmann & Jusczyk, 1998). However, we found that even 12-month-old infants can detect novel nonadjacent relationships if they can bootstrap from simpler instances of such structure (but not from exposure to nonstructural similarities such as vocabulary and prosody), underscoring the importance of considering prior experience in theories of language acquisition.

English-learning infants begin to show sensitivity to syntactic categories in their language by 14 months (Waxman & Booth, 2001). Consistent with previous work (Gerken et al., 2005; Gómez & LaKusta, 2004), we found that infants can use distributional and phonological information to learn co-occurrence dependencies between word categories (e.g., that *a* words combine with *X*s but not *Y*s). Critically, we showed that 12-month-old infants are also beginning to generalize from such adjacent dependencies to more difficult nonadjacent ones, despite the fact that the adjacent and nonadjacent dependencies occurred in strings with dissimilar surface features and that the strings themselves were novel. Generalizing across such surface features is critical to learning language's abstract syntactic patterns.

Why, however, should females benefit over males? Given that females' verbal memory tends to exceed males' across development (e.g., Kramer, Delis, Kaplan, O'Donnell, & Prifitera, 1997), one possibility is that females have better associative memory for phonologically related words. If females form more cohesive groups of *X* and *Y* elements based on the syllable–number cue, they may be better able to generalize to novel nonadjacent *aX* and *bY* combinations. Hartshorne and Ullman (2006) found that female toddlers are more likely to generalize the past-tense ending to irregular verbs (e.g., saying *holded* instead of *held*), particularly for irregular verbs sharing phonological features with regular verbs (e.g., the irregular verb *hold* is phonologically similar to the regular verbs *fold* and *mold*). Hartshorne and Ullman suggested that females formed stronger associations between phonologically related verbs and that they inappropriately generalized the regular past-tense morphology to irregular phonological neighbors as a result. Whether these findings reflect differences in memory development or other processes involved in language acquisition,[1] they pose an intriguing question for further research on sex differences in language development.

These findings speak to the role of experience in language acquisition. Experience is an undeniable factor in this process (infants exposed to English learn English, not French); however, its specific role has long been debated. On one account, infants are constrained to

[1]Females lead males in word learning (Nelson, 1973) and early word combinations (Schacter, Shore, Hodapp, Chalfin, & Bundy, 1978), and all of these sex differences may be driven by differences in neural development.

represent syntactic structure in adult like ways (Crain, 1991; Pinker, 1984; Wexler, 1990), with changes in language competence reflecting the addition of new information rather than changes in the representation of such structures themselves. In contrast to this view, our findings add to a growing literature suggesting that infants can track statistical structure relevant to language within the first year of life, and that such learning develops through experience.

## Acknowledgments

This research was supported by National Institutes of Health Grant R01 HD42170. We thank the parents and infants who participated and Merrill Garrett, LouAnn Gerken, Jessica Payne, Mary Peterson, and three anonymous reviewers for comments on earlier drafts.

## References

Chomsky, N. Syntactic structures. The Hague, The Netherlands: Mouton; 1957.

Cohen, LB.; Atkinson, DJ.; Chaput, HH. Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0) [Computer software]. Austin: University of Texas; 2004.

Crain S. Language acquisition in the absence of experience [Target article plus commentaries]. Behavioural and Brain Sciences. 1991; 4:597–650.

Elman JL. Learning and development in neural networks: The importance of starting small. Cognition. 1993; 48:71–99. [PubMed: 8403835]

Fisher C. The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). Cognition. 2002; 82:259–278. [PubMed: 11747864]

Gerken LA, Wilson R, Lewis W. Seventeen-month-olds can use distributional cues to form syntactic categories. Journal of Child Language. 2005; 32:249–268. [PubMed: 16045250]

Gertner Y, Fisher C, Eisengart J. Learning words and rules: Abstract knowledge of word order in early sentence comprehension. Psychological Science. 2006; 17:684–691. [PubMed: 16913951]

Gómez RL. Variability and detection of invariant structure. Psychological Science. 2002; 13:431–436. [PubMed: 12219809]

Gómez RL, Gerken LA. Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. Cognition. 1999; 70:109–135. [PubMed: 10349760]

Gómez RL, LaKusta L. A first step in form-based category abstraction in 12-month-old infants. Developmental Science. 2004; 7:567–580. [PubMed: 15603290]

Gómez RL, Maye J. The developmental trajectory of nonadjacent dependency learning. Infancy. 2005; 7:183–206.

Hartshorne JK, Ullman MT. Why girls say "holded" more than boys. Developmental Science. 2006; 9:21–32. [PubMed: 16445392]

Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. Cognition. 2002; 83:B35–B42. [PubMed: 11869728]

Kramer JH, Delis DC, Kaplan E, O'Donnell L, Prifitera A. Developmental sex differences in verbal learning. Neuropsychology. 1997; 11:577–584. [PubMed: 9345701]

Lany J, Gómez RL, Gerken LA. The role of prior experience in language acquisition. Cognitive Science. 2007; 31:481–507.

Monaghan P, Chater N, Christiansen MH. The differential role of phonological and distributional cues in grammatical categorization. Cognition. 2005; 96:143–182. [PubMed: 15925574]

Nelson K. Structure and strategy in learning to talk. Monographs of the Society for Research in Child Development. 1973; 38:1–135.

Newport EL. Maturational constraints on language learning. Cognitive Science. 1990; 14:11–28.

Newport EL, Aslin RN. Learning at a distance: I. Statistical learning of non-adjacent dependencies. Cognitive Psychology. 2004; 48:127–162. [PubMed: 14732409]

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

Rhode DLT, Plaut DC. Language acquisition in the absence of explicit negative evidence: How important is starting small? Cognition. 1999; 72:67–109. [PubMed: 10520565]

Saffran JR, Aslin RN, Newport EL. Statistical learning by eight-month-old infants. Science. 1996; 274:1926–1928. [PubMed: 8943209]

Saffran JR, Thiessen ED. Pattern induction by infant language learners. Developmental Psychology. 2003; 39:484–494. [PubMed: 12760517]

Saffran JR, Wilson DP. From syllables to syntax: Multilevel statistical learning by 12-month-old infants. Infancy. 2003; 4:273–284.

Santelmann LM, Jusczyk PW. Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. Cognition. 1998; 69:105–134. [PubMed: 9894402]

Schacter FF, Shore E, Hodapp R, Chalfin S, Bundy C. Do girls talk earlier? Mean length of utterance in toddlers. Developmental Psychology. 1978; 14:388–392.

Smith LB, Jones SS, Landau B, Gershkoff-Stowe L, Samuelson L. Object name learning provides on-the-job training for attention. Psychological Science. 2002; 13:13–19. [PubMed: 11892773]

Thiessen ED, Saffran JR. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. Developmental Psychology. 2003; 39:706–716. [PubMed: 12859124]

Thiessen ED, Saffran JR. Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. Language Learning and Development. 2007; 3:73–100.

Tomasello M. Do young children have adult syntactic competence? Cognition. 2000; 74:209–253. [PubMed: 10640571]

Waxman SR, Booth AE. Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. Cognitive Psychology. 2001; 43:217–242. [PubMed: 11689022]

Wexler K. Innateness and maturation in linguistic development. Developmental Psychobiology. 1990; 23:645–660. [PubMed: 2286296]

**Fig. 1.**

Diagram of the three phases of the experimental procedure, showing the language materials from Version 1. During familiarization, infants heard two-word phrases, each consisting of a word from category *a* or *b* and a word from category *X* or *Y*. The experimental and control groups differed in the specific words that were paired, as indicated in the figure. Phrases in italics were withheld from familiarization. The subset of withheld strings that were grammatical to both the experimental and the control groups was used in novel nonadjacent dependencies during habituation. These dependencies were violated in the test strings. Version 2 contained the opposite pairings during familiarization, and thus, the strings presented at habituation and test were swapped (e.g., infants were habituated to the phrases *ong hes deech* and *alt hes coomo* in Version 2 and were tested on the phrases *ong hes coomo* and *alt hes deech*).

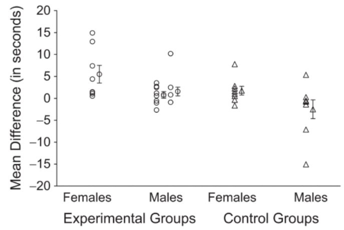


**Fig. 2.**

Male and female infants' mean increase in listening times from the last two habituation trials to the two test trials in the experimental and control groups. Each group's mean and standard error are depicted just to the right of the column containing individuals' values. The data for the males in the experimental group are divided, with the individual and mean values for the original 8 infants to the left and the data for the additional 4 individuals and the mean for all 12 infants to the right.

## TABLE 1

Habituation Measures by Familiarization Condition and Sex

| Group | Number of looks | Accumulated listening time (s) |
|---|---|---|
| Experimental | | |
| Female | 10.4 (1.5) | 115.6 (37.7) |
| Male | 7.9 (1.7) | 118.3 (26.4) |
| Control | | |
| Female | 6.6 (1.6) | 86.9 (20.9) |
| Male | 9.9 (1.7) | 165.97 (38.1) |

**Note.** Standard errors are shown in parentheses. There were no significant differences across condition and sex for either total number of looks or accumulated listening time, $Fs(1, 31) \leq 1.76$, $ps \geq .18$. In addition, neither habituation measure was a significant predictor of discrimination across groups when entered into a regression, $\beta s \leq .29$, $ps \geq .13$. The model including both measures did not significantly predict performance, $R^2 = .095$, $F(2, 29) = 1.53$, $p = .23$.