Research Article

VARIABILITY AND DETECTION OF INVARIANT STRUCTURE

Rebecca L. Gómez

The Johns Hopkins University

Abstract—*Two experiments investigated learning of nonadjacent dependencies by adults and 18-month-olds. Each learner was exposed to three-element strings (e.g., pel-kicey-jic) produced by one of two artificial languages. Both languages contained the same adjacent dependencies, so learners could distinguish the languages only by acquiring dependencies between the first and third elements (the nonadjacent dependencies). The size of the pool from which the middle elements were drawn was systematically varied to investigate whether increasing variability (in the form of decreasing predictability between adjacent elements) would lead to better detection of nonadjacent dependencies. Infants and adults acquired nonadjacent dependencies only when adjacent dependencies were least predictable. The results point to conditions that might lead learners to focus on nonadjacent versus adjacent dependencies and are important for suggesting how learning might be dynamically guided by statistical structure.*

Recent research has led to exciting discoveries regarding early learning abilities (Gómez & Gerken, 1999; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; see Gómez & Gerken, 2000, for a review). For example, 8-month-olds can use the higher transitional probabilities of adjacent syllables within words versus the lower transitional probabilities of syllables spanning words to identify wordlike units in continuous speech (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). Such learning appears to be rapid, involuntary, and domain general. It occurs after as little as 2 min of exposure (Saffran, Aslin, & Newport, 1996) and when learners are not explicitly focused on or aware of statistical structure in the training stimuli (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; see also Gómez, 1997); it also appears to be age invariant, occurring as readily for infants as for young children and adults (Saffran et al., 1997). Learners can use transitional probabilities to segment rapid streams of tones (Saffran, Johnson, Aslin, & Newport, 1999) as well as temporal sequences of visual shapes (Fiser & Aslin, 2002). Such learning also occurs in nonhuman primates, and thus is not limited to humans (Hauser, Newport, & Aslin, 2001). The ubiquitous, robust, and fundamental nature of this learning suggests that learners are particularly sensitive to adjacent dependencies.

One might like to know how learners acquire nonadjacent (or remote) dependencies, in addition to adjacent dependencies. Nonadjacent dependencies seem to pose a considerable challenge by requiring learners to form relations over irrelevant intervening material. Examples found in language are dependencies between auxiliaries and inflectional morphemes (*e.g., is running, has eaten*), as well as dependencies involving number agreement (*The rocks on the bluff are jagged*). Learners also have to track remote dependencies in high-level cognitive tasks involving event knowledge (eating dinner or getting ready for work), means-ends analysis, and planning.

One possible basis for such learning is that higher-order conditionals build on lower-level ones such that first-order conditionals are embedded in knowledge of second-order ones (where first-order conditionals require knowledge of the immediately preceding element and second-order conditionals require knowledge of the previous two elements). According to this view, learners chunk adjacent dependencies and then form a higher-order relation between the chunked pair and the next element in sequence. Such learning has been studied extensively using the serial reaction time (SRT) task, in which learners track the location of a cursor on a visual display by pressing a corresponding response key (e.g., Cohen, Ivry, & Keele, 1990). The cursor follows a predetermined sequence. Learning is indicated by faster reaction times to trials consistent with the sequence than to trials disrupting the sequential structure. Although sequences with higher-order dependencies are more difficult to acquire than those with lower-order dependencies, learners are clearly able to track both kinds of structure (Cleeremans & McClelland, 1991; Curran & Keele, 1993; Reed & Johnson, 1994; Schvaneveldt & Gómez, 1998).

Although some sequences are predictable (highly ritualized tasks like tying one's shoes or eating at a fast-food restaurant), others are considerably more variable. In language, for instance, certain elements belong to relatively small sets (function morphemes like *a, was, -s,* and *-ing*), whereas others belong to very large sets (nouns and verbs). Additionally, small and large sets are interspersed because the functional elements in small sets associate syntactically with elements in larger sets (e.g., verbs pattern with auxiliaries like *was* and inflections like *-ing*). This patterning translates into lower transitional probabilities for *YIX* because this probability decreases as the set size of *Y* increases. The learning required for event knowledge and high-level planning is similar when certain events or procedures are fairly fixed and thus belong to small sets whereas others are more variable (consider how certain steps in meal preparation are fixed whereas others vary depending on the menu). The elements from the smaller sets figure prominently in many remote dependencies, and so a learner able to track these elements is at a considerable advantage compared with one who cannot.

The decreased transitional probabilities accompanying set-size differences may present a problem for learning involving higher-order dependencies because decreases in predictability are known to lead to segmentation of sequential stimuli (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). However, perhaps learners are able to capitalize on set-size differences such that elements from small sets are perceived as more or less stable as a function of the variability of their surrounding context. If so, perhaps decreases in transitional probabilities actually cause learners to switch their focus to other sources of invariant structure. In that case, learners might track remote dependencies more easily when they involve elements that are stable in comparison to their changing context.

My students and I examined this possibility by exposing participants to one of two artificial languages. Each language produced nonsense strings like *pel-wadim-jic* or *vot-kicey-rud*. The strings in the two languages began and ended with the same words and contained the same pair-wise transitions. They differed, however, in the dependencies between their first and third elements. Thus, learners could distinguish

---

Address correspondence to Rebecca L. Gómez, Department of Psychology, University of Arizona, Tucson, AZ 85721-0068; e-mail: rgomez@u.arizona.edu.

the languages only by acquiring the nonadjacent dependencies. We also manipulated the variability of context by systematically increasing the size of the pool from which we drew the middle element.

This design allowed us to test two competing hypotheses. Learners embedding lower-order dependencies in higher-order ones should learn nonadjacent dependencies better on exposure to small sets than on exposure to larger sets because transitional probabilities between adjacent elements are higher for smaller than for larger set sizes. In contrast, if increasing variability leads learners to focus on other sources of invariant structure, then learners should show rising sensitivity to nonadjacent dependencies when set size increases. We investigated these competing hypotheses in two experiments with adults and infants.

EXPERIMENT 1

Method

Participants

Forty-eight undergraduate psychology students at Johns Hopkins University participated for extra credit.

Stimuli

During training, each participant listened to auditory strings produced by one of two artificial languages (L1 or L2, shown in Fig. 1a).[1] L1 strings took the form *aXd*, *bXe*, and *cXf*. L2 strings took the form *aXe*, *bXf*, and *cXd*. Variability was manipulated by drawing X from a pool of either 2, 6, 12, or 24 elements.

The elements *a*, *b*, and *c* were instantiated as *pel*, *vot*, and *dak*; *d*, *e*, and *f* were instantiated as *rud*, *jic*, and *tood*. The 24 X elements were *wadim, kicey, puser, fengle, coomo, loga, gople, taspu, hiftam, deecha, vamey, skiger, benez, gensim, feenam, laeljeen, chila, roosa, plizet, balip, malsig, suleb, nilbo,* and *wiffle*. The set of 12 X elements consisted of the first 12 words in the list, the set of 6 consisted of the first 6, and the set of 2 consisted of the first 2. Six strings in each language were common to all four sets. These were used as test stimuli (see Table 1). Thus, six of the test strings were generated by L1, and six were generated by L2.

A female speaker recorded sample strings. Word tokens from the recorded strings were used to construct both L1 and L2 strings. Tokens were used to eliminate talker-induced differences in individual strings. There were 250-ms pauses between words and 750-ms pauses between strings, so that words were distinguishable in strings and strings were distinguishable from one another.

Procedure

Six subjects participated in each of the eight conditions resulting from the language (L1 vs. L2) and set-size (2, 6, 12, 24) manipulations.

During training, learners exposed to set size 24 heard six iterations of each of 72 (3 dependencies × 24 X elements) training strings. They were thus exposed to a total of 432 strings. Frequency of exposure to
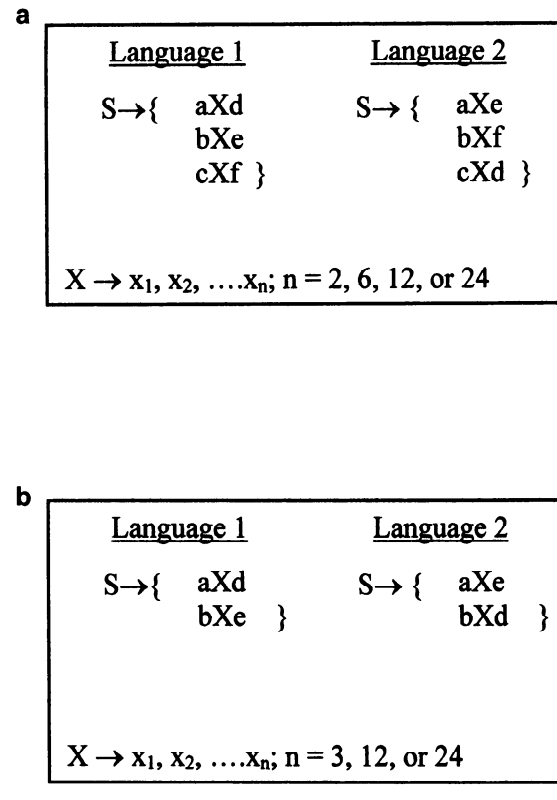

Fig. 1. Structure of the languages used in Experiment 1 with adult participants (a) and in Experiment 2 with 18-month-old infants (b).

the nonadjacent dependencies was held constant across set size by exposing participants in the other set-size conditions to 432 strings as well (e.g., participants exposed to set size 12 encountered each string twice as often as those exposed to set size 24).

Training lasted approximately 18 min. Learners were told to listen to the novel language for a subsequent test. Before the test, they were told that the strings they had heard during training were generated according to a set of rules involving word order, and they would now hear 12 strings, 6 of which they had heard previously. Half of the strings would follow the same word order as the training strings, and half would

Table 1. *Test strings used in Experiment 1 (identified by language)*

Language 1	Language 2
pel wadim rud	pel wadim jic
vot wadim jic	vot wadim tood
dak wadim tood	dak wadim rud
pel kicey rud	pel kicey jic
vot kicey jic	vot kicey tood
dak kicey tood	dak kicey rud

Note. The order of strings was randomized for each participant. Note that the discriminations required were very subtle.

violate the order. They were instructed to select the "Y" key on a keyboard if they thought a string followed the rules determining word order and to select the "N" key otherwise.

Results and Discussion

Table 2 presents the percentage of endorsements for trained versus untrained strings in each of the four set-size conditions. Note that the differences in strings were extremely subtle, for example, requiring learners to distinguish *pel-wadim-rud* from *pel-wadim-jic*.

An analysis of variance with set size and materials (L1 vs. L2) as between-subjects variables and grammaticality (trained vs. untrained strings) as a within-subjects variable resulted in a main effect of grammaticality, $F(1, 40) = 45.30, p < .001$, but more important, a significant Grammaticality × Set Size interaction, $F(3, 40) = 4.757, p < .01$. There were no other main effects or interactions. Most notably, comparisons between adjacent set-size conditions revealed a significant increase in performance between set size 12 and set size 24, $t(22) = 2.82, p = .009$, but not between set sizes 2 and 6 or set sizes 6 and 12, $ts ≤ 0.77, ps ≥ .208$. Additional tests comparing endorsement rates for trained versus untrained strings within conditions were highly variable for the smaller set sizes: marginal for set size 2, $t(11) = 1.84, p = .092$, and set size 12, $t(11) = 2.19, p = .051$, but significant for set size 6, $t(11) = 2.82, p = .016$. The test for set size 24 was also significant, $t(11) = 7.31, p ≤ .001$. There were also individual differences in learning. A very small number of participants exposed to the smaller set sizes learned the nonadjacent dependencies so well as to discriminate the trained and untrained strings perfectly (2 of 12 participants in set sizes 2 and 6, 3 of 12 in set size 12); however, a markedly greater proportion of learners (8 of 12) showed perfect discrimination in set size 24.

There was little evidence that learners were embedding first-order dependencies in higher-order ones. If they had, performance would have been significant for set size 2 and would have decreased with increasing set size. Neither was there evidence for gradually improved performance with increasing variability of context. Rather, performance increased radically for the very largest set size. Recall that both languages contained the same pair-wise transitions. Thus, learners focused primarily on adjacent dependencies should have been unable to distinguish trained and untrained test strings. In fact, performance was only modest for the smaller set-size conditions (accuracy was 60.5%, 66.5%, and 65.5% for set sizes 2, 6, and 12, respectively).[2] The relatively modest performance for the smaller set sizes in combination with the radical increase for the largest set size (90% accuracy) suggests that learners may have focused on adjacent dependencies as the default, switching to nonadjacent dependencies only when the former were sufficiently unreliable. In order to avoid confusion regarding rule-governed recognition versus old/new recognition, it is important to point out that although participants were asked to make grammaticality judgments, all discriminations were for old versus new strings.[3] Nevertheless, such findings are important for demonstrating the role of statistical structure in guiding learning, as well as the remarkable adaptability of human learners.

Although the findings suggest that increasing variability of context leads to better learning of nonadjacent dependencies by adults, one might

---

2. The slightly above-chance performance in the small set-size conditions results from the small number of participants showing perfect discrimination and raises interesting questions regarding individual differences in learning.

3. We have subsequently obtained similar results in a replication with old/new judgments.

Table 2. *Percentage endorsements for trained and untrained strings in Experiment 1*

	Nonadjacent dependency				Difference (trained − untrained)	
	Trained		Untrained			
Set size	M	(SEM)	M	(SEM)	M	(SEM)
2	88	(6)	68	(10)	20	(11)
6	87	(6)	54	(11)	33	(12)
12	77	(7)	47	(10)	30	(14)
24	100	(0)	20	(11)	80	(11)

also like to know how such variability affects younger learners. Previous research has shown that infants can distinguish nonadjacent dependencies in natural language by 18 months of age, but not earlier (Santelmann & Jusczyk, 1998). We decided to test this age group to see whether their learning mirrored that of the adults.

EXPERIMENT 2

Method

Participants

Forty-eight infants (24 males, 24 females) with an average age of 18 months 18 days (range: 17 months 11 days to 19 months 18 days) were tested. Eighteen other infants were tested but not included because of excessive fussiness (12) or technical difficulties (6).

Stimuli

The stimuli were identical to those used in Experiment 1, except that infants were trained on two rather than three nonadjacent dependencies (see Fig. 1b).

Each infant was trained on one of two artificial languages with a set size of 3, 12, or 24. The beginning, ending, and middle elements were identical to those used in Experiment 1. Infants exposed to set size 24 heard each of the 48 strings (2 dependencies × 24 X elements) produced by their training language one time during training. As with the adults, we equated frequency of exposure to the nonadjacent dependencies across conditions. Therefore, infants exposed to set size 12 heard each of the 24 (2 × 12) training strings twice, whereas infants exposed to set size 3 heard each of the 6 (2 × 3) training strings eight times. Training lasted approximately 3 min.

Six strings in each language were used in all three set-size conditions (the subset of strings used for the set size 3 condition). These were used as test stimuli (see Table 3). Each test trial consisted of the set of six strings from one of the languages, in one of two random orders. The four test sets (L1 in Order 1, L1 in Order 2, L2 in Order 1, L2 in Order 2) were each presented twice during the test (once in each of two test blocks), for a total of eight test trials. Test sets were 17 s in duration.

Procedure

Each infant was tested individually while seated on the caregiver's lap in an enclosed booth, using the head-turn preference procedure

Table 3. *Test strings used with 18-month-olds (Experiment 2)*

Language 1	Language 2
pel wadim rud	pel wadim jic
vot wadim jic	vot wadim rud
pel kicey rud	pel kicey jic
vot kicey jic	vot kicey rud
pel puser rud	pel puser jic
vot puser jic	vot puser rud

Note. All six strings for a particular language were presented during each test trial, in randomized order. Note that the discriminations required were extremely subtle.

Table 4. *Mean listening time (in seconds) for infants tested in Experiment 2*

	Nonadjacent dependency				Difference (trained − untrained)	
	Trained		Untrained			
Set size	*M*	*(SEM)*	*M*	*(SEM)*	*M*	*(SEM)*
3	6.74	(0.86)	6.79	(0.94)	−0.04	(0.57)
12	7.46	(0.87)	7.12	(0.46)	0.35	(0.83)
24	' 5.23	(0.45)	7.30	(0.48)	−2.07	(0.40)

(see Kemler Nelson et al., 1995). An observer outside the test booth monitored the infant's looking behavior using a button box connected to an Apple Powermac. The experimental control program initiated trials and scored head-turn responses. To eliminate bias, both the caregiver and the observer listened to masking stimuli over headphones. During training, stimuli were presented simultaneously from two loudspeakers located on either side of the infant. The infant's gaze was directed first toward a blinking middle light and then toward one of two blinking side lights (one below each loudspeaker). When the infant looked away from the side light for 2 s, his or her gaze was again directed toward the middle. There was no relationship between lights and sound during training.

During the test, each trial began with the center light blinking. Once the infant fixated on the center light, the experimenter pressed a button to extinguish it. This action initiated blinking of one of the side lights (the one associated with the source of sound for that trial). When the infant turned his or her head in the direction of the side light by 30°, the test set for that trial played until the infant looked away for 2 s (or until the trial played out). The observer recorded the direction of the infant's head turns. The computer program tracked looking times, the amount of time looking away from the source of sound (terminating trials after 2 s), and controlled the randomization and presentation of stimuli. The dependent measure was the amount of time an infant oriented toward the test stimulus. A significant difference in listening time to trained versus untrained strings would indicate that the infants acquired some sensitivity to the nonadjacent dependencies defined by their training language.

Results and Discussion

Mean listening times are shown in Table 4. An analysis of variance with set size (3 vs. 12 vs. 24) and materials (L1 vs. L2) as between-subjects variables and grammaticality (trained vs. untrained strings) as a within-subjects variable resulted in a significant Grammaticality × Set Size interaction, $F(2, 42) = 4.026, p \leq .05$. There were no other main effects or interactions. Fifteen of the 16 infants exposed to set size 24 showed a novelty effect, listening significantly longer to untrained than to trained strings, $t(15) = 5.11, p = .0001$. This significant difference in looking times suggests that they were able to discriminate the two stimulus types. Infants exposed to set sizes 3 and 12 showed no discrimination whatsoever ($ts \leq 0.42, ps \geq .680$). Half of the infants in each of these conditions showed a novelty effect. Additional comparisons between adjacent set-size conditions revealed greater

listening time differences for infants exposed to set size 24 versus set size 12, $t(30) = 2.62, p = .014$, but not for those exposed to set size 12 versus set size 3, $t(30) = 0.39, p = .700$.

As with the adult learners, there was no evidence that the infants were embedding first-order dependencies in higher-order ones. Nor was there evidence for a gradual increase in discrimination with increasing set size. Rather, discrimination increased radically for the largest set size. Such learning was all the more remarkable given the short exposure times in this study (3 min). Additionally, the discrimination required of infants was extremely subtle (e.g., *pel-kicey-rud* vs. *pel-kicey-jic* and *vot-wadim-jic* vs. *vot-wadim-rud*). The failure to discriminate on exposure to the smaller set sizes in combination with the abrupt increase in discrimination for the largest set size suggests that the infants, like the adults, processed adjacent dependencies as the default, and switched their focus to nonadjacent dependencies only when the former were sufficiently unreliable. The findings are important for demonstrating that very young learners can respond to the informational demands of their learning situation and indicate a fair degree of flexibility in their learning.

GENERAL DISCUSSION

We used a miniature artificial language to investigate learning involving nonadjacent dependencies. Remote dependencies are critical to the extent that learners must coordinate elements separated in sequence. We incorporated set-size differences like those found in naturally occurring sequences, such that some elements are fixed while others belong to much larger sets and, hence, are more variable. Set-size differences affect the predictability of transitions between adjacent elements and, thus, should affect the resulting learning. Learners might take one of two approaches to tracking remote dependencies: (a) They could embed lower-order transitions in higher-order conditionals, or (b) high variability of context might lead them to focus on the less variable structure, namely, the nonadjacent dependencies.

Learning involving the formation of higher-order conditionals should have been reflected in better performance for smaller than for larger set sizes because the transitional probabilities were higher for smaller set sizes. However, discrimination was poor in these conditions. Instead, discrimination increased markedly under conditions of greatest variability, suggesting that learners are sensitive to change versus nonchange (Gibson, 1991) and use this sensitivity to home in on stable structure. The pattern of findings was the same for infants and adults.

Learners also appear to adapt to the informational demands of their input. For instance, the tendency to focus on adjacent dependencies ap-

pears to be fundamental in learning and may even be the default. But what happens when such dependencies are no longer reliable, as they often are not in sequential structure? The present studies suggest that learners will focus on a particular source of information only to the extent that it yields some degree of statistical regularity. Beyond this point, learners seek out alternative sources of predictability. The ability to focus attention on one source of information over another, depending on the particular characteristics of the input, suggests a fair degree of flexibility in learning.

It is noteworthy that learners show such flexibility in the absence of explicit corrective feedback, given that such feedback is thought to play a critical role in language acquisition. This is informative because of earlier assumptions that statistical learners would be ill equipped to distinguish relevant from irrelevant correlations in the input (e.g., Pinker, 1984).

A few additional points bear mention. Although the results suggest that transitional probabilities may be subject to a minimum threshold of reliability, there is no reason to believe that 24 is a magic number. For example, slightly younger infants can acquire nonadjacent dependencies with a set size of 18 (Gómez & Maye, 2001). Rather, there seems to be some point, depending on the statistical structure of the learning environment, at which the remote dependencies are perceived as invariant. Of critical interest will be a more thorough investigation of the interplay of processes involved in such learning. In particular, do learners engage multiple learning processes (e.g., one focusing on adjacent and another on nonadjacent dependencies), or is the same general process operative under a wide array of conditions? These findings would seem to raise intriguing questions for models of sequential learning. Additionally, although performance in the small set-size conditions was poor for most participants, a small number of learners in these conditions showed perfect discrimination, raising interesting questions regarding individual differences in learning. It is also important to acknowledge that the argument that learners focused on adjacent dependencies in the small set-size conditions is at present based on indirect evidence. Studies investigating direct evidence for this claim are currently under way.

Finally, the poor discrimination in the low-variability conditions was somewhat surprising given that exposure to individual strings was so high (a consequence of holding exposure to the nonadjacent dependencies constant across conditions). Although more exposure should result in better learning, perhaps participants were overexposed to the materials. Poor discrimination may have reflected boredom or some other form of decreased sensitivity. We investigated this possibility in a follow-up study by holding exposure to individual strings constant. Using the strings from Experiment 1, we exposed adult learners (with set sizes of 2, 6, and 12) to six iterations only of each string produced by their training language (matching the number of iterations for learners in the set size 24 condition in Experiment 1). The endorsement rates for trained versus untrained strings are shown in Table 5. Note that the means for set size 24 are taken from Experiment 1.

Although performance was better than chance for learners exposed to set size 12, it was marginal at best for set sizes 2 and 6 and did not differ significantly between adjacent set-size conditions for set sizes 2, 6, and 12. There was still an abrupt increase in performance between set sizes 12 and 24. Thus, it does not appear that the low discrimination rates observed in Experiment 1 can be explained by overexposure during training. Interestingly, Newport and Aslin (2000) also failed to find learning of nonadjacent dependencies with materials similar to those used in the set size 2 condition. In contrast to the participants in the present studies, their learners faced the added difficulty of having to acquire remote dependencies in continuous speech.

Table 5. *Percentage endorsements for trained and untrained strings in the follow-up study*

	Nonadjacent dependency				Difference (trained – untrained)	
	Trained		Untrained			
Set size	*M*	*(SEM)*	*M*	*(SEM)*	*M*	*(SEM)*
2	81	(6)	78	(10)	3	(8)
6	80	(6)	69	(6)	11	(5)
12	80	(9)	44	(12)	36	(14)
24	100	(0)	20	(11)	80	(11)

Note. The means for set size 24 are from Experiment 1.

In conclusion, recent research suggests that human learners are extremely sensitive to statistical structure (Aslin et al., 1998; Cleeremans & McClelland, 1991; Gómez & Gerken, 1999; Hasher & Zacks, 1984; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Schvaneveldt & Gómez, 1998). The present studies add to these findings by demonstrating how learning might be dynamically guided by such structure. Although it is not yet clear whether statistical learning reflects one general process or multiple specialized processes, it does appear to be subject to one unifying principle, namely, a tendency toward "reduction of uncertainty" (Gibson, 1991). Learners are driven to seek invariant structure in the stimulus array. When transitional probabilities are high, adjacent elements will be perceived as invariant. When high variability disrupts adjacent dependencies, learners will seek alternative sources of predictability. The present findings are consistent with the view that humans are active, opportunistic learners, ready to capitalize on whatever regularities might come their way.

REFERENCES

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.

Braine, M.D.S. (1965). The insufficiency of a finite state model for verbal reconstructive memory. *Psychonomic Science, 2,* 291–292.

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

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

Curran, T., & Keele, S.W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 189–202.

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

Gibson, E.J. (1991). *An odyssey in learning and perception.* Cambridge, MA: MIT Press.

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

Gómez, R.L., & Gerken, L.A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

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

Gómez, R.L., & Maye, J. (2001). *Developmental trends in acquiring non-adjacent dependencies*. Unpublished manuscript, University of Arizona, Tucson.

Hasher, L., & Zacks, R.T. (1984). Automatic processing of fundamental information. *American Psychologist, 39*, 1372–1388.

Hauser, M.D., Newport, E.L., & Aslin, R.N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*, B53–B64.

Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., & Gerken, L.A. (1995). The Head-turn Preference Procedure for testing auditory perception. *Infant Behavior and Development, 18*, 111–116.

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

Newport, E.L., & Aslin, R.N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S.C. Howell, S.A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 1–21). Somerville, MA: Cascadilla Press.

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

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determin-

ing what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 585–594.

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

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

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.

Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101–105.

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

Schvaneveldt, R.W., & Gómez, R.L. (1998). Attention and probabilistic sequence learning. *Psychological Research, 61*, 175–190.