



Grammatical pattern learning by human infants and cotton-top tamarin monkeys

Jenny Saffran ^{a,*}, Marc Hauser ^{b,c}, Rebecca Seibel ^a,
Joshua Kapfhamer ^a, Fritz Tsao ^d, Fiery Cushman ^b

^a *Department of Psychology & Waisman Center, University of Wisconsin-Madison,
1202 West Johnson Street, Madison, WI 53706-1696, USA*

^b *Department of Psychology, Harvard University, 980 William James Hall,
33 Kirkland Street, Cambridge, MA 02138, USA*

^c *Department of Organismic & Evolutionary Biology, Harvard University, Cambridge,
MA 02138, USA*

^d *Department of Psychology, Hillsdale College, Hillsdale, MI 49242, USA*

Received 18 September 2006; revised 4 September 2007; accepted 28 October 2007

Abstract

There is a surprising degree of overlapping structure evident across the languages of the world. One factor leading to cross-linguistic similarities may be constraints on human learning abilities. Linguistic structures that are easier for infants to learn should predominate in human languages. If correct, then (a) human infants should more readily acquire structures that are consistent with the form of natural language, whereas (b) non-human primates' patterns of learning should be less tightly linked to the structure of human languages. Prior experiments have not directly compared laboratory-based learning of grammatical structures by human infants and non-human primates, especially under comparable testing conditions and with similar materials. Five experiments with 12-month-old human infants and adult cotton-top tamarin monkeys addressed these predictions, employing comparable methods (familiarization–discrimination) and materials. Infants rapidly acquired complex grammatical structures by using statistically predictive patterns, failing to learn structures that lacked such patterns. In contrast, the tamarins only exploited predictive patterns when

* Corresponding author. Tel.: +1 608 262 9942; fax: +1 608 262 4029.
E-mail address: jsaffran@wisc.edu (J. Saffran).

learning relatively simple grammatical structures. Infant learning abilities may serve both to facilitate natural language acquisition and to impose constraints on the structure of human languages.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Infants; Monkeys; Grammar learning; Statistical learning

1. Introduction

Human language is a vastly complex system that is nevertheless mastered early in life. The question of how children discover structures that are not transparently mirrored in the input is central not only to investigations of how language acquisition proceeds, but also to our fundamental understanding of the nature of human language. One such feature of human language is phrase structure. Phrases frequently contain unidirectional dependency relations: the presence of some word categories requires others, which conjointly make up a phrase [e.g., nouns can occur without articles like *the* and *a*, but an article rarely occurs without a noun somewhere downstream]. These predictive dependency relationships provide a statistical cue that highlights phrasal units for learners (Saffran, 2001).

Why might languages consistently include predictive dependencies within phrases? One possibility is that predictive dependencies enhance learnability; the presence of these structural relationships may help learners to discover phrases. Previous studies tested this hypothesis by comparing the acquisition of two artificial languages (Saffran, 2002). The Predictive language included predictive dependencies as a cue to phrase structure (e.g., A-words could occur with or without D-words, but D-words required A-words). The Non-predictive language lacked predictive dependencies (e.g., A-words could occur with or without D-words and vice versa; neither word class predicted the presence of the other). Adults and school-aged children exposed to the Predictive language outperformed those exposed to the Non-predictive language on subsequent tests of language structure. Similar results were obtained using sequences created from non-linguistic materials, suggesting that this particular constraint on learning is not specifically tied to language (Saffran, 2002). These findings are consistent with results obtained across many distinct literatures suggesting that predictability facilitates the discovery of underlying structure.

A critical question then is: what does a learning ability that is not restricted to language learning have to do with language at all? Is the ability to detect predictive dependencies so prevalent that its relationship to linguistic structure is accidental? That is, perhaps languages use predictive dependencies, and human learners use predictive dependencies, but these two facts are independent and provide no mutual explanatory power. Alternatively, predictive dependencies may be part of human languages *because* they enhance learnability.

A growing body of research suggests that languages exemplify exactly those structures that humans are best able to learn (e.g., Christiansen & Chater,

1999; Christiansen & Dale, 2004; Morgan, Meier, & Newport, 1987; Morgan & Newport, 1981; Newport & Aslin, 2000). This perspective contrasts with earlier construals of Universal Grammar, in which linguistic structure was specified by innate knowledge. Instead, at least some aspects of structure may emerge from constraints imposed by learning itself. On this view, the relationship between a particular learning mechanism and language structures is not accidental. To the extent that human languages have been sculpted by constraints imposed by learning, we would expect to see a close match between the kinds of mechanisms deployed by learners and the structures that typify languages. However, despite the inherent interest in this hypothesis, it has not been tested with infant learners (though see Saffran & Thiessen, 2003; for an example from phonology). If it is the case that linguistic structure and learning mechanisms are intimately related, it is necessary to test learners who are themselves at the beginning stages of language acquisition, in order to determine whether the constraints on learning in question are operating at the necessary point in development.

The present study thus had three goals. First, we explored whether infants could successfully acquire grammars, ranging in surface complexity, which included predictive dependencies akin to those found in natural languages. Second, we assessed the related hypothesis that infants would fail to acquire highly similar grammars that contained structures that are unlike natural languages. Third, we systematically contrasted human infant and tamarin monkey performance on these languages to begin an investigation of potential differences in the learning mechanisms available across species. We note here that comparing only two species, with only one method, is insufficient to make any broad phylogenetic claims; we see this work as a starting point intended to invite a broader taxonomic and methodological exploration (Hauser, Barner, & O'Donnell, 2007).

We began our investigation by asking whether infants can make use of predictive dependencies in learning an artificial grammar. To do so, we adapted the languages used by Saffran (2002) for use with infant participants. Infants were exposed to one of two languages: a *Predictive (P)-language*, in which predictive dependencies marked phrase units, and a *Non-predictive (NP)-language*, that lacked predictive dependencies. While both languages contained elements that predicted one another, only the P-language contained predictive elements specifically placed within phrases. This dependency relationship provided a cue to phrasal units in the P-language, which was not available in the NP-language. Because prior infant studies have typically used language systems that generate far fewer possible sentences than the languages used by Saffran (2002) with adult and child participants, we decided to begin our infant studies by generating the sentence exemplars using only one word token from each word class. While this manipulation greatly simplified the materials presented during the learning phase relative to the languages used by Saffran (2002) (in which word classes contained multiple word tokens), these materials still maintained the critical structural distinction between the P-language and the NP-language: the presence or absence of predictive dependencies within phrases.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Twenty-four infants (12.5–13.0 months) were randomly assigned to each of the two experimental conditions. All infants were full-term, monolingual, and free of ear infections. Half of the infants (mean age: 12.8 months) were exposed to the P-language and half (mean age: 12.9 months) were exposed to the NP-language.

2.1.2. Materials

The materials consisted of sentences generated from two artificial grammars. Each element in the grammars represents an item from a nonsense-word class (see Table 1).

The **P-Language** contained predictive dependencies between word classes:

$$S \rightarrow AP + BP + (CP)$$

$$AP \rightarrow A + (D)$$

$$BP \rightarrow CP + F$$

$$CP \rightarrow C + (G)$$

S refers to Sentence, AP refers to A-phrase, BP to B-phrase, etc. Parentheses denote optional elements. For example, the possible A-phrases are A and AD. Thus, if a D-word is present it must be preceded by an A-word; however, the presence of an A word does not predict the presence of a D word. The predictive patterns are thus uni-

Table 1
Word Classes used in the Artificial Grammars

1A: Experiments 1/3 Word Classes

A	biff
C	cav
D	klor
F	dupp
G	Jux

1B: Experiment 2/4 Word Classes

A	biff	hep	mib	rud
C	cav	lum	neb	sig
D	klor	pell		
F	dupp	loke	jux	vot
G	tiz	pilk		

1C: Experiment 5 Word Classes

A	biff	hep
C	cav	lum
D	Klor	pell
G	Tiz	jux

directional. C- and G-words follow a similar within-phrase dependency. The conditional probability for within-phrase elements (left to right) is 1.0 [e.g., $D|A = 1.0$]. The directionality of these dependencies (backwards) is the opposite of English (forwards – the predictive items, e.g. determiners, precede the elements they predict, e.g. nouns).

The **NP-Language** lacked predictive dependencies between word classes:

$S \rightarrow AP + BP$

$AP \rightarrow \{(A) + (D)\}$ [must contain at least one]

$BP \rightarrow CP + F$

$CP \rightarrow \{(C) + (G)\}$ [must contain at least one]

Due to the optionality of so many elements, dependencies within phrases were not present. For example, the possible A-phrases are A, D, and AD. In this case, A words and D words do not predict one another, as each can occur in the absence of the other. This grammar does have a phrase structure of a sort; there is a negative predictive relationship between same-phrase elements: if A is *not* present, then D must be present, and vice versa. This structure is uncharacteristic of natural languages.

Familiarization. Sentence exemplars were generated using one token from each word class (Table 1A). Eight sentences from each of the P- and NP-languages were recorded by a trained female speaker using descending sentential prosody, with 1 s of silence between sentences.

Test. Four sentences shared between the P- and NP-languages served as grammatical test items, and four sentences that were not possible in either language served as ungrammatical test items. The same test items were used to test all infants, regardless of exposure language, to control for idiosyncratic item preferences. Due to the small size of the language, grammatical test sentences were drawn from the familiarization corpus (Table 2A). Ungrammatical sentences were novel for all infants.

2.1.3. Procedure

Infants were familiarized for 3 min while playing quietly with their parent in a sound-attenuated booth. Infants then received a 2 min refamiliarization session designed to provide exposure to the lights used during testing. During refamiliarization, the lights were flashed contingent on the infant's looking behavior. The subsequent test employed the Headturn Preference Procedure, in which we measured the duration of headturns towards concealed audio-speakers. Each test item consisted of two sentences in alternation, separated by a 1-s period of silence. Half of the items were grammatical, the other half were ungrammatical. A test trial began with a flashing light at the center of the wall facing the infant. When the experimenter signaled the computer that the infant had fixated on the center light, one of the side lights began to flash, and the center light was extinguished. As soon as the infant made a head turn in the direction of the flashing side light, one of the test items was presented from the speaker beneath the light. When the infant looked away for more

Table 2
Familiarization and Test Strings

2A. Experiments 1/3

P-Language Familiarization Strings

biff cav dupp
biff cav jux dupp
biff klor cav jux dupp
biff klor cav dupp
biff cav dupp cav jux
biff klor cav dupp cav jux
biff cav dupp cav
biff klor cav dupp cav

Grammatical Test Strings

biff cav dupp
biff klor cav jux dupp
biff klor cav dupp
biff cav jux dupp

NP-Language Familiarization Strings

biff cav dupp
biff cav jux dupp
biff klor cav jux dupp
biff klor cav dupp
klor cav jux dupp
biff klor jux dupp
klor cav dupp
biff jux dupp

Ungrammatical Test Strings

biff cav klor dupp
biff jux cav dupp
biff klor dupp
biff cav jux klor dupp

2B. Experiments 2/4 (The 50 familiarization strings for each of the two languages used in Experiments 2/4 are provided in the appendices of Saffran (2002))

Grammatical Test Strings

hep pell cav pilk dupp
hep klor cav tiz vot
mib klor cav jux
rud pell neb pilk loke

Ungrammatical Test Strings

hep pilk cav pell dupp
rud pilk lum klor vot
hep klor cav tiz
bif pilk lum klor jux

2C. Experiment 5

Familiarization Strings

biff cav
biff lum
hep cav
hep lum
hep klor cav
biff pell cav
hep pell cav
biff klor lum
hep klor lum
biff pell lum
hep pell lum
hep cav tiz
biff lum tiz
hep lum tiz
biff cav jux
hep cav jux

biff lum jux
hep lum jux
biff klor cav tiz
hep klor cav tiz
hep pell cav tiz
biff klor lum tiz
hep klor lum tiz
biff pell lum tiz
hep pell lum tiz
biff klor cav jux
hep klor cav jux
biff pell cav jux
hep pell cav jux
biff klor lum jux
biff pell lum jux
hep pell lum jux

Hard Grammatical Test Strings

biff klor cav
biff pell cav tiz^a
hep klor lum jux

Hard Ungrammatical Test Strings

hep klor tiz
biff jux cav
hep lum klor jux
biff cav pell tiz

Easy Grammatical Test Strings

biff cav jux
hep pell lum

Easy Ungrammatical Test Strings

lum pell biff
tiz cav hep

Table 2 (continued)

hep klor cav jux	jux cav klor hep
biff klor lum tiz	tiz lum pell biff
hep lum tiz ^a	

^a Items containing these two sentences were not included in the analyses due to an error in the design of the test materials.

than 2 s, the test item stopped playing, and the center light began to flash again. This procedure continued until infants completed the 12 test trials (three trials for each of the four test items, presented in random order).

2.2. Results

The infants showed significant discrimination between grammatical and ungrammatical sentences after exposure to the P-language, but not the NP-language (see Fig. 1). In the P-Language condition, infants listened significantly longer to ungrammatical items than to grammatical items [$t(11) = 2.52, p < .05$]. In the NP-Language condition, there was no statistically significant difference in listening times [$t(11) = .23, p = .8$]. An ANOVA examining the interaction between items (grammatical and ungrammatical) and conditions (P-language and NP language) was not significant [$F(1,22) < 1, n.s.$], likely due to variance in the NP-language condition.

The fact that only infants exposed to the P-language showed a significant test discrimination supports the hypothesis that 12-month-old infants can make use of predictive dependencies in grammars written over individual words. Their corresponding failure to learn the highly similar NP-language materials is striking given the small size of this language (just eight sentences), and the use of grammatical test items that were actually played during familiarization. Furthermore, there was

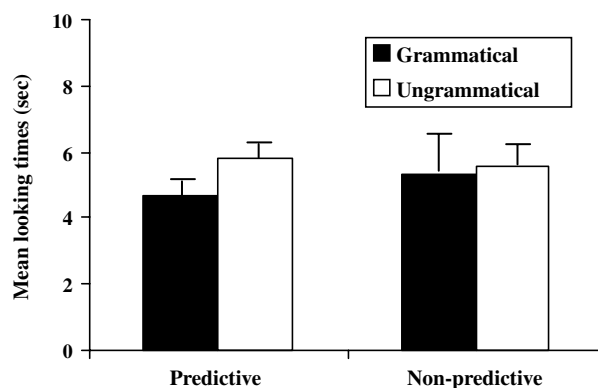


Fig. 1. Experiment 1 results. Mean (SE) infant looking times to grammatical and ungrammatical test tokens for the P-language and NP-language.

no evidence that the failure to learn the NP-language was due to decreased attention when contrasted with the P-language: the average duration of the infants' test looks after exposure to the P-language (5.25 s) and the NP-language (5.46 s) were comparable.

We next explored the hypothesis that this pattern of results is due to detection of the kinds of low-level statistical patterns investigated in prior infant research (Aslin et al., 1998; Gómez & Gerken, 1999; Saffran, Aslin, & Newport, 1996) as well as extensive adult research (e.g., Perruchet & Pacteau, 1990; Reber, 1989). To do so, we analyzed the experimental materials, computing the average word frequencies, bigram (word-pair) frequencies and probabilities (forwards and backwards), and trigram frequencies for the grammatical and ungrammatical test items relative to the P-language and NP-language familiarization corpora. There was no evidence that the P-language (Table 3A) and NP-language (Table 3B) familiarization corpora differed in these particular n-gram properties relative to the test items; the ratios between the statistics for the grammatical and ungrammatical test items were equivalent across the two exposure language corpora.

These analyses do not rule out the possibility that other low-level statistical patterns contributed to the learning outcomes. Any given set of artificial language materials invariably contains myriad statistical patterns beyond those that are explicitly controlled or analyzed (e.g., Seidenberg, MacDonald, & Saffran, 2002), and relatively little is known about which patterns are detected by infants – other regularities may have contributed to these results. Moreover, infant experimental designs typically employ relatively few test stimuli, analyzed en masse, making it difficult if not impossible to track the effects of very specific cues. For example, consider the ungrammatical test sequence **biff-cav-klor-dupp*. On the one hand, this sequence might be rejected by Language P learners based on the *cav-klor* bigram if they expect *cav-dupp* instead, or based on the *klor-dupp* bigram if they expect

Table 3

Average n-gram statistics for the Experiment 1 test materials relative to the exposure sentences; frequencies computed with one pass through the exposure corpus

	Grammatical items	Ungrammatical items	Ratio (%)
<i>3A. Statistics computed relative to the P-Language exposure corpus</i>			
Unigram frequency	8.00	7.50	93
Bigram frequency	4.13	1.71	41
Trigram frequency	2.5	0.13	5
Forwards bigram probability	0.67	0.41	61
Backwards bigram probability	0.60	0.30	50
<i>3B. Statistics computed relative to the NP-Language exposure corpus</i>			
Unigram frequency	6.25	6.11	97
Bigram frequency	3.23	1.35	42
Trigram frequency	1.75	0.13	7
Forwards bigram probability	0.65	0.34	52
Backwards bigram probability	0.52	0.21	40

cav-dupp instead, as compared to NP learners.¹ On the other hand, one might predict that **biff-cav-klor-dupp* should have been preferred by P learners over NP learners because the only familiar bigram, *biff-cav*, occurred twice as often in the P language as the NP language. There are numerous such features, across both languages, which learners could have exploited in principle. The important point is that while these low level regularities occurred throughout both the P language and NP language exposure corpora, infants were more successful on the P language materials. The presence of the predictive dependencies in the P-language may have facilitated detection of those patterns that are most closely aligned to the structure of these languages.

Notably, the languages we used in Experiment 1 were vastly simpler than the structures found in natural languages along many dimensions. One important difference is that the Experiment 1 materials involved patterns over individual words, rather than word classes or types. In Experiment 2, we again contrasted the P-language and the NP-language, but this time we used a larger vocabulary, with lexical items grouped into categories (A, C, etc.). Experiment 2 thus used the full languages from Saffran (2002), with sentence exemplars generated using multiple tokens from each word class. These materials thus more closely approximate the operation of predictive dependencies in natural languages.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Twenty-four infants (12.5–13.0 months) were randomly assigned to each of the two experimental conditions. All infants were full-term, monolingual, and free of ear infections. Half of the infants (mean age: 12.7 months) were exposed to the P-language and half (mean age: 12.7 months) were exposed to the NP-language.

3.1.2. Materials

The languages were identical to Experiment 1, except that sentence exemplars were generated using multiple tokens from each word class (Table 1B). Note that in both languages, there were no phonological cues to word classes (A words, D words, etc). The A, C, and F classes each contained four words, while the D and G classes contained two words. This asymmetry was meant to reflect class size differences between content and function words in natural languages; function words, typically the predictor items given within-phrase predictive dependencies, are generally smaller categories than the content word categories that they predict. In order to discover these classes, learners need to track the distributions of individual words (e.g., which words tend to precede and follow) and to group words with overlapping

¹ We thank an anonymous reviewer for providing this example.

distributions into categories, as hypothesized for natural language word classes (e.g., Mintz, Newport, & Bever, 2002; Mintz, 2002). The vocabulary consisted of 16 words, arrayed into 50 unique sentences during familiarization. Unlike Experiment 1, the grammatical and ungrammatical test items were all novel sentences not presented during familiarization (Table 2B).

3.1.3. Procedure

The procedure was identical to Experiment 1, except that the infants received 23-min of familiarization in a booth not used for testing. We increased this exposure because the 5 min exposure period from Experiment 1 seemed too brief to learn something as complex as the Experiment 2 materials. Parents were instructed to encourage snacking and quiet play. After familiarization, infants were moved to the test booth, and received a brief refamiliarization (30 s) with the artificial language materials, with lights contingent on eye gaze, prior to testing.

3.2. Results

The pattern of results was identical to Experiment 1: infants in the P-Language condition listened significantly longer to ungrammatical items than to grammatical items [$t(11) = 2.77, p < .05$], while infants in the NP-Language condition did not show a statistically significant difference [$t(11) = .002, p = .99$] (see Fig. 2). An ANOVA examining the interaction between items (grammatical and ungrammatical) and conditions (P-language and NP language) was significant [$F(1,22) = 6.4, p < .05$].

These results are striking even when placed in the context of previous work pinpointing powerful infant learning abilities (for recent reviews, see Gómez, 2006; Saffran et al., 2006). In the P-language condition, infants successfully acquired knowledge about the distribution of words in a novel language far more complex than those used in prior experiments; the structure of this language can generate several thousand unique sentences. At the same time, infants failed to show evidence of

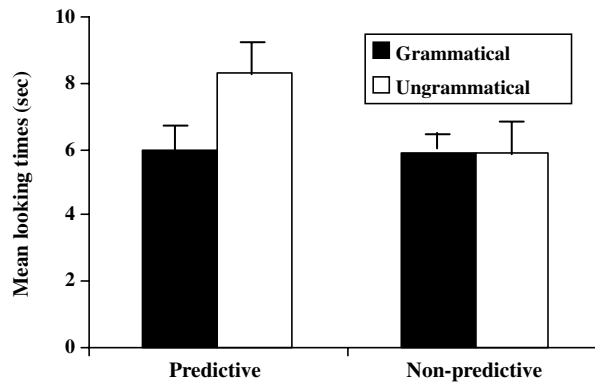


Fig. 2. Experiment 2 results. Mean (SE) infant looking times to grammatical and ungrammatical test tokens for the P-language and NP-language.

learning in the NP-language condition, despite the general structural similarity of these two language systems. These findings, in tandem with Experiment 1, suggest that predictive dependencies may in fact facilitate learning even for infants. The observed pattern of successes and failures for infant learners is consistent with the hypothesis that patterns that are observed widely across the languages of the world are more learnable by infants than those that are not. Notably, unlike adult studies of this type, the current results are not vulnerable to the criticism that learners are simply imposing structures from their own native language onto the artificial grammars; 12-month-old infants are unlikely to have acquired much in the way of native language grammatical structure. These results thus provide support for the claim that some grammatical patterns are more easily learned than others, and that cross-linguistic regularities may be shaped by infant learning mechanisms.

An alternative argument is that perhaps the predictive languages are *inherently* more learnable, by any type of learner, and that success on such grammars is unrelated to the learner's ability to acquire natural languages. One way to examine this hypothesis is to contrast humans with non-human learners. In some cases, we see apparently isomorphic learning capacities. For example, human infants and cotton-top tamarin monkeys are both facile at spontaneously (i.e., in the absence of training or significant exposure) tracking the transitional probabilities between syllables (Saffran et al., 1996; Hauser, Newport, & Aslin, 2001), and can discover and generalize simple rule-like sequences in syllable strings (Marcus, Vijayan, Rao, & Vishton, 1999; Hauser, Weiss, & Marcus, 2002). In other cases, however, there are divergences. For example, whereas both human adults and tamarins compute non-adjacent transitional probabilities, only human performance maps on to natural language structure (Newport & Aslin, 2004; Newport, Hauser, Spaepen, & Aslin, 2004). Similarly, whereas human adults spontaneously acquire both finite state [e.g., (AB)ⁿ] and phrase structure [e.g., AⁿBⁿ] grammars, tamarins only spontaneously acquire the former (Fitch & Hauser, 2004). These results build on prior work suggesting that the computational abilities needed to acquire a phrase structure grammar [hierarchical patterning, embedding] may not be spontaneously available to some non-human animals (e.g., Hauser, Chomsky, & Fitch, 2002; Premack & Premack, 2002; Savage-Rumbaugh, 1986), although they are able to learn about some kinds of non-linguistic hierarchical structure such as cup seriation and kinship organization (e.g., Frigaszy, Galloway, Johnson-Pynn, & Brakke, 2002; Greenfield, 1991; Bergman, Beehner, Cheney, & Seyfarth, 2003; McGonigle, Chalmers, & Dickinson, 2003), and with significant training, at least one species – the European starling – can acquire AⁿBⁿ (Gentner, Fenn, Margoliash, & Nusbaum, 2006). As noted by a number of commentators (e.g., Hauser et al., 2007; Perruchet & Rey, 2005), however, acquisition of AⁿBⁿ can be successfully accomplished by means of multiple mechanisms.

Experiments 3 and 4 were thus designed to replicate Experiments 1 and 2 using a non-human primate: the cotton-top tamarin. We followed prior research in using protocols that are comparable across these two species. Because previous studies have not directly compared human infants and tamarins on complex grammar learning tasks (the closest being the ABA/ABB tasks used by Marcus et al., 1999 with infants and Hauser et al., 2002), Experiment 3 used the relatively simple grammar

from Experiment 1, with the P-language and the NP-language written over individual word tokens. Experiment 3 was designed to ask whether tamarins can learn systems of this complexity under conditions comparable to the infants and whether, like infants, the tamarins can exploit predictive dependencies.

4. Experiment 3

4.1. Methods

4.1.1. Participants

Twenty-three adult cotton-top tamarins (*Saguinus oedipus*), 13 males and 10 females participated in Experiment 3. All animals were born in captivity, live in social groups consisting of a mated pair and their offspring, and have extensive experience in laboratory tasks. Each tamarin was familiarized and then tested on both the P- and NP-languages, with approximately two weeks between testing (order counter-balanced). Seven participants in the P-Language condition and 4 in the NP-Language condition were not included in the analyses due to difficulties moving them to the test apparatus or poor behavior prior to/during test.

4.1.2. Materials

Identical to Experiment 1.

4.1.3. Procedure

We used a familiarization-test procedure deployed in prior experiments with this colony (Fitch & Hauser, 2004; Hauser et al., 2001, 2002; Newport et al., 2004). The evening before testing, we presented the tamarins with either the P- or NP-language for 2 h in their home-room. The next day, each subject received 2-min of refamiliarization followed immediately by testing. Data were coded as in prior experiments by measuring an orienting response to a test stimulus presented from a concealed loud-speaker while the subject was still and facing away from the speaker [<http://www.wjh.harvard.edu/~mnkylab/prothome.htm>.] Subjects were scored as responding if they turned to look toward the speaker either during the presentation of the test stimulus, or within 2 s afterwards. Responses were not reinforced. Trials consisted of four grammatical and ungrammatical items with the constraint that a given trial not follow a previous trial within 10-s and beyond 60-s. Order of grammatical and ungrammatical test items was counterbalanced. Inter-observer reliabilities (blind coding) ranged between 0.85 and 0.90.

4.2. Results

Like the infants, the tamarins showed significant discrimination between grammatical and ungrammatical sentences after exposure to the P-language, but not the NP-language (see Fig. 3). In the P-Language condition, tamarins responded significantly more to the ungrammatical items than the grammatical items [Wilcoxon,

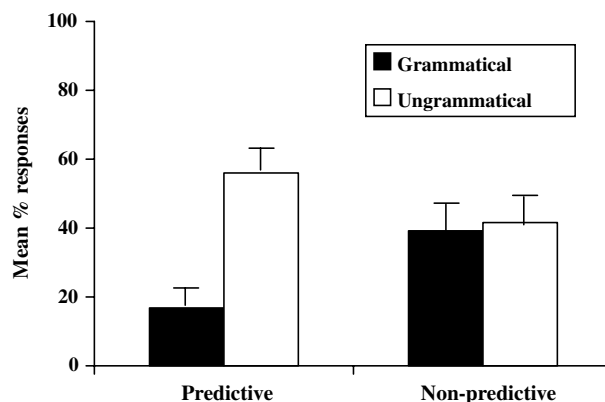


Fig. 3. Experiment 3 results. Mean [SE] proportion of responses made by tamarins to grammatical and ungrammatical test tokens for the P-language and NP-language.

$z = 3.19$, $p = .001$]. In contrast, they showed no statistically significant pattern of response to the NP language [Wilcoxon, $z = 0.4$, $p = .69$]. An ANOVA examining the interaction between items (grammatical and ungrammatical) and conditions (P-language and NP language) was significant [$F(1, 33) = 4.56$, $p < .05$].

These results suggest that tamarins are able to detect regularities in a simple grammar, written over individual word tokens, when predictive dependencies are present. However, they fail to make the same test discrimination following exposure to stimuli lacking predictive dependencies. As with the infants, this failure to discriminate the test sentences following exposure to the NP-language is surprising because the grammatical test sentences were played during exposure, and thus should be relatively easy to remember. Given stimuli of this limited complexity, the two species appear to perform in parallel. We thus ran Experiment 4, which was an attempt to replicate the pattern of infant successes (P-language) and failures (NP-language) using the richer set of stimuli from Experiment 2.

5. Experiment 4

5.1. Methods

5.1.1. Participants

The same set of 23 tamarins from Experiment 3 participated, 10 months after completing Experiment 3. As in Experiment 3, each tamarin was exposed to and tested on both the P- and NP-languages, with approximately two weeks between successive runs of the P- and NP-languages (order counterbalanced). All tamarins successfully completed this task. Importantly, we expected little to no carry over from the prior experiment given the lack of carry over effects in previous research where different conditions are run on the same animals, often with less time in between conditions (e.g., Newport et al., 2004).

5.1.2. Materials

Identical to Experiment 2.

5.1.3. Procedure

Identical to Experiment 3.

5.2. Results

The tamarins maintained the same overall level of responses as in Experiment 3 (33% for the P-language, and 29% for the NP-language). However, tamarins did not show a statistically significant pattern of response in either the P-Language condition [Wilcoxon, $z = -.21$, $p = .84$] or the NP-Language condition [Wilcoxon, $z = -1.57$, $p = .12$] (see Fig. 4). The tamarins failed to discriminate between grammatical and ungrammatical strings in Experiment 4 even when predictive dependencies were present. This pattern of results diverges from the infants in Experiment 2, who showed significant discrimination between grammatical and ungrammatical sentences after exposure to these P-language materials, but not the NP-language language materials.

Why did the tamarins fail to learn the predictive language in Experiment 4, when they succeeded in Experiment 3? One possibility is that tamarins, and presumably other non-human animals, are able to spontaneously learn links between individual elements (as in finite-state grammars), but fail to learn relationships between categories of elements (as in phrase structure grammars) in the absence of explicit training (Fitch & Hauser, 2004; Hauser et al., 2002; Gentner et al., 2006; for review of these issues, see Hauser et al., 2007). However, it is unclear whether the infants in Experiment 2 themselves learned categories in this fashion. The materials did not test generalization to new category members. Deriving a test that would do so raises a methodological challenge, given that there were no phonological regularities or reference words to mark the categories. Thus, while it is possible that the infants in

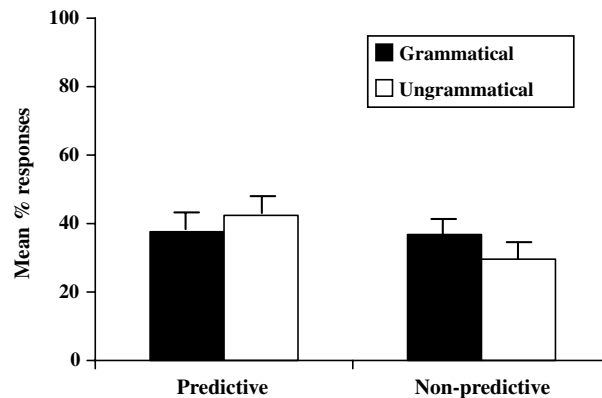


Fig. 4. Experiment 4 results. Mean [SE] proportion of responses made by tamarins to grammatical and ungrammatical test tokens for the P-language and NP-language.

Experiment 2 did learn the distributionally-defined categories and the patterns linking them into phrases, there is reason to believe that they may instead have learned token-level patterns. A large body of literature in adults demonstrates the challenge of learning categories given only distributional information (e.g., (Braine, 1987; Frigo & McDonald, 1998; Smith, 1969; though see Mintz, 2003; for a counterexample). The extant artificial language literature on linguistic category induction in infancy includes categories in which distributional cues are correlated with phonological cues (e.g., Gerken et al., 2005; Gómez & LaKusta, 2004). It is thus unclear whether there are any circumstances in which infants can discover categories in artificial languages from purely distributional information, in the absence of other correlated cues; this question is the object of current research.

Returning to the tamarins, their failure to show any evidence of learning in Experiment 4 could thus be accounted for by information-processing challenges. In particular, the Experiment 4 materials are more complex than those in Experiment 3 across a number of dimensions, which likely increased the task demands inherent in these experiments substantially. For example, the Experiment 4 lexicon was three times the size of the Experiment 3 lexicon, increasing the memory load. Similarly, the tamarins may have failed to generalize beyond the sentences presented during familiarization, which would have affected performance in Experiment 4 [although these same animals have been shown to generalize over simpler materials in experiments using similar methods (Hauser et al., 2002)].

We designed Experiment 5 to begin to disentangle potential causes for the tamarins' failure in Experiment 4. We hypothesized that some factors were more likely than others to affect the tamarins' performance relative to the infants. For example, since tamarins responded at the same general level in Experiments 3 and 4, but only discriminated the test items following exposure to the P-language of Experiment 3, such factors as motivation and the distinctiveness of the materials seemed unlikely to explain the observed pattern of responses. A more likely explanation is that infants are better able to track and remember large numbers of elements in sequence than tamarins, perhaps especially when the material is speech – which is far more familiar to the infants than to the tamarins. To test this possibility, the simplified grammar used in Experiment 5 employed the predictive dependencies from the prior P-languages, but consisted of fewer and shorter sentence types and tokens than Experiment 4 (see Table 4).

We also elaborated our repertoire of test items. In particular, within a test session we presented both “Easy” and “Hard” test contrasts. Easy Grammatical items

Table 4
Stimulus comparison: Experiments 1/3 (P-language), Experiments 2/4 (P-language) and Experiment 5

	Experiments 1/3	Experiments 2/4	Experiment 5
# of sentence types	8	12	4
# of sentence tokens in exposure	8	50	32
Mean sentence length (words)	4.5	4.24	3.375
Vocabulary size	5	16	8

consisted of strings heard during familiarization, whereas Easy Ungrammatical items consisted of highly novel reverse-ordered strings. In contrast, Hard Grammatical items consisted of novel grammatical strings whereas Hard Ungrammatical items consisted of single sentence-internal violations, like those tested in Experiments 1–4. Successful discrimination of the Hard test items would suggest that tamarins *can* use predictive dependencies in linguistic environments more complex than Experiment 3 – though less complex than Experiment 4. However, a failure in Experiment 5 would suggest that the tamarins’ difficulties in Experiment 4 were unlikely to be due to the size of the lexicon or the length of the sentences. The inclusion of the Easy test items allowed us to ask whether tamarins could learn even simple aspects of these stimuli. A failure to learn these items, accompanied by levels of response comparable to prior experiments, would indicate that even relatively simple pattern learning is easily disrupted in the tamarins. We also included infants as a control group in order to ensure that these materials are in fact learnable, and to provide an additional replication of the infant results from the P-language conditions of Experiments 1 and 2.

6. Experiment 5

6.1. Methods

6.1.1. Participants

Twelve 12.5-month-old infants (mean age = 12.8 months), and 20 tamarins (from the same group as the previous experiments) participated in this experiment. Three tamarins were excluded from the analyses due to poor behavior during the test.

6.1.2. Materials

The grammar for Experiment 5 contained predictive dependencies and was intermediate in complexity between the P-languages of Experiments 1/3 and 2/4:

$$S \rightarrow AP + CP$$

$$AP \rightarrow A + (D)$$

$$CP \rightarrow C + (G)$$

The language was generated using two tokens from each word class (Table 1C), with 36 possible sentences (see Table 4 for cross-experiment comparisons). The grammatical structure was simpler than the Experiments 2/4 P-language (no F word or sentence-final optional CP). When considering patterns of word types, this language is simpler than the P-language of Experiments 1/3 (just 4 sentence types, as opposed to the 8 types from Experiment 1/3). However, in terms of patterns of word tokens, it is more complex than the P-language from Experiments 1/3 (36 potential sentences in Experiment 5 versus 8 in Experiments 1/3).

Five of the eight grammatical test sentences (Easy Grammatical items) were heard during familiarization (as in Experiments 1/3; Table 2C). The other three (Hard

Grammatical items) were novel (as in Experiments 2/4).² Four of the eight ungrammatical sentences (Easy Ungrammatical) contained words in reverse order and thus contained multiple violations (e.g., *GCA). The other four (Hard Ungrammatical), were similar to the ungrammatical sentences used in Experiments 1–4: they contained legal beginning and ending elements and violated sentence-internal phrase structure rules (e.g., *ADG). As in the prior experiments, each test item consisted of two sentences.

6.1.3. Procedure

Identical to Experiment 2/4, except that the infants received 6.5-min of familiarization.

6.2. Results

Tamarins showed no statistically significant differences in their overall pattern of responses to grammatical and ungrammatical strings [Wilcoxon, $z = 0.41$, $p = .68$; see Fig. 5A].³ The tamarins failed to discriminate the Hard Ungrammatical and Hard Grammatical items [Wilcoxon, $z = .32$, $p = .74$]. They also failed to discriminate the Easy Ungrammatical and Easy Grammatical items [Wilcoxon, $z = .85$, $p = .40$]. The tamarins' overall level of response to the test materials was 29%, comparable to Experiments 3 and 4. These results thus replicate the tamarins' failure to learn the P-language grammar in Experiment 4.

Infants listened significantly longer overall to ungrammatical than to grammatical items [$t(11) = 5.94$, $p < .001$; see Fig. 5B]. The infants successfully discriminated between Easy Grammatical and Easy Ungrammatical items [$t(11) = 5.07$, $p < .001$]. They also discriminated between Hard Grammatical and Hard Ungrammatical items [$t(11) = 4.40$, $p < .01$]. As in Experiments 1 and 2, infants appear to have had no difficulty acquiring the predictive language presented in Experiment 5.

7. General discussion

The results of these experiments provide a clear illustration of the power of infant learning mechanisms. While it is of course evident that infants rapidly acquire the rich structural properties of their *native* language over the first two years of life, these studies are the first to show rapid acquisition of grammatical patterns in laboratory learning tasks of this complexity. After a few minutes of exposure to the P-language, infants in Experiment 1 were able to discriminate grammatical from ungrammatical sentences in a simple language environment in which the relevant patterns involved the distribution of individual word tokens. With a little more exposure, infants in

² An error in stimulus design led to the imbalance in the numbers of Easy and Hard Grammatical test sentences.

³ One test item was excluded from the reported analyses because it contained both an Easy and a Hard Grammatical sentence. Including this item in the analyses does not change the pattern of results.

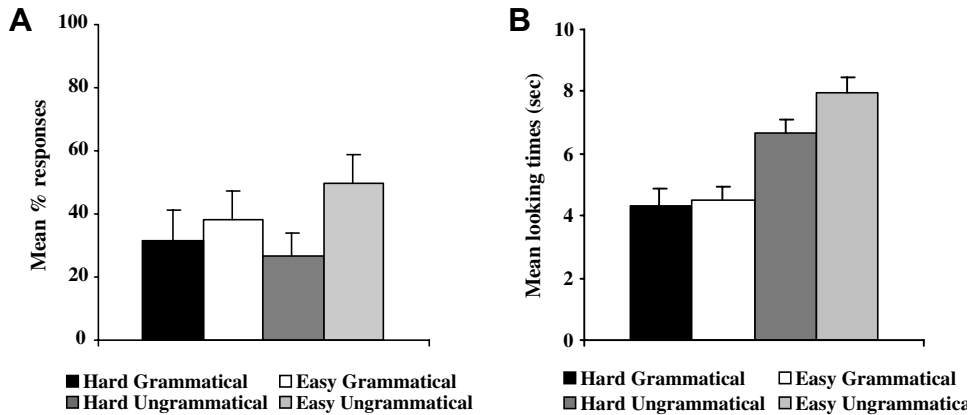


Fig. 5. Experiment 5 results. Mean [SE] proportion of responses made by tamarins (A) and infants (B) in response to Hard Grammatical, Easy Grammatical, Hard Ungrammatical, and Easy Ungrammatical test tokens.

Experiment 2 were able to discriminate grammatical from ungrammatical sentences in a far more complex P-language, one that generates several thousand possible sentences. The successes on the P-languages presented in Experiments 1, 2 and 5 are particularly striking given the age of these infants; these 12-month-olds are only just beginning to produce their first words. These findings broadly support the claim emerging from numerous artificial and natural language studies that infants' linguistic knowledge vastly outstrips language production.

Of potentially greater significance than these successes on the P-languages were infants' corresponding failures to perform the same test discriminations following exposure to the NP-languages. In Experiment 1, the infants did not discriminate the test sentences despite the relative simplicity of this language. Similarly, in Experiment 2, we found no evidence that the infants were able to acquire the NP-language. The primary difference between the P-languages and the NP-languages was the presence or absence of predictive dependencies linking words within phrases together. That said, it is never possible to rule out all other distributional differences between the languages, and it is always possible that the infants would have learned the NP-languages given additional exposure. However, the consistent finding that only those languages with predictive dependencies were learned in this paradigm does support the claim that infants can make use of strong links between words to begin to bracket the input into chunks, facilitating learning.

Turning to the tamarin data, the results suggest that tamarins can learn *some* grammatical patterns. In particular, they detected predictive dependencies between individual elements in Experiment 3; they appear to be capable of attending to these kinds of stimuli and are motivated to respond during testing. Notably, the tamarins failed to exploit the predictive patterns in Experiments 4 and 5, even though their overall level of responsiveness was the same as in Experiment 3, suggesting that

attention and motivation to respond are unlikely to account for the pattern of results.⁴

At some level, these results might be considered unsurprising. After all, human infants learn human languages, and cotton-top tamarins do not. Similarly, it seems obvious that humans can learn more complex materials than cotton-top tamarins. Our interest in these findings resides in the nuanced pattern of results. Infants do not always learn more than tamarins. For example, both infants and tamarins successfully performed the test discrimination following exposure to the P-language in Experiments 1 and 3, while both failed to perform the test discrimination following exposure to the NP-language. Similarly, both groups failed to perform the test discrimination in Experiments 2 and 4 following exposure to the NP-language. It is thus not simply the case that infants learn more complex structures than tamarins. The structure of the materials to be learned critically determines the outcome of learning.

There are of course many possible differences between infants and tamarins that may influence learning performance, including retention and learning rate. We tested adult tamarins with, by definition, more mature brains than human infants. However, the infants possess human cognitive skills (at least in the fledgling phase, en route to the adult form). This crossing of developmental state and species makes it difficult to predict a priori which factors are most likely to help or hinder learning in these two groups. The tamarins received the bulk of their exposure to the languages the night before testing, unlike the infants, who were familiarized with the materials just prior to testing. This delay may have led to forgetting on the part of the tamarins, though they did receive a refamiliarization just prior to testing; in prior work with this species, this is precisely the method used, which in many cases yielded strong evidence of learning. Alternatively, the overnight delay may have assisted memory consolidation; in fact, there is suggestive evidence that tamarins show better test performance given an overnight delay compared to exposure just prior to testing (Hauser, unpublished data). Similarly, recent results with human infants suggest that sleep may help consolidation during language learning, particularly given tasks that require generalization from a set of input (Gómez, Bootzin, & Nadel, 2006).

It is also possible that the tamarins may require more exposure in order to learn. In the current study, the tamarins received markedly more exposure than the infants: 2 h in each experiment, versus 5 min (Experiment 1), 23.5 min (Experiment 2), and 7 min (Experiment 5) for the infants. Of course, it is possible that had we increased the tamarin exposure five-fold across experiments, as done for the infants, better learning would have occurred. But the point remains that the tamarins were unable to learn as well as the infants even with substantially more exposure. These results are also limited by the inclusion of only a single non-human species, and need to be extended to include additional species [e.g., other primates, songbirds with

⁴ A few weeks after the conclusion of Exp. 5, the same animals were tested in a replication of the Hauser et al. (2001) word segmentation study. The tamarins showed successful discrimination, taking advantage of transitional probabilities to segment a continuous stream of speech syllables. This suggests that the animals had not habituated in general, and nor had they lost motivation to respond to artificial language materials.

complex hierarchically organized learned vocal signals] and more complex linguistic structures [e.g., recursive operations, phrase structures with more intervening material between predictive elements]. While the implementation of widely used experimental protocols is helpful for discerning convergent evidence, it is also important to explore the consequences of different comparative methods [e.g., operant training procedures]. The current study represents only the beginning of a research program designed to contrast human infants with non-human animals, which will necessarily require parametric variation of the many potentially important variables not assessed in the present experiments (Hauser et al., 2007).

Moving beyond species-specificity, an interesting related question raised by these studies concerns the domain-specificity of this type of learning. In studies with adults using the P-language and NP-language, Saffran (2002) found that predictive dependencies assisted adult learners in non-linguistic tasks designed to be analogous to the language-learning task. Thus, for example, when learners were presented with sequences of non-linguistic sounds (computer alert sounds in one study, or drums and bells in another), or simultaneously-presented visual arrays of non-linguistic shapes, the P-language learners outperformed the NP-language learners. These findings led Saffran (2002, 2003) to speculate that constraints to cluster the input based on predictive dependencies are not limited to language, and might be observed across a number of domains. While the current studies are limited to linguistic stimuli, it will be of great interest to ask whether infants continue to show enhanced learning when predictive dependencies are instantiated in non-linguistic materials.

To the extent that human learning abilities have shaped the structure of natural languages, our results provide a potential explanation for two central facts about human language that are often viewed as inconsistent: extensive cross-linguistic similarities and powerful human learning abilities. If humans are such good learners, why are languages so similar? Our findings are consistent with the hypothesis that natural languages have been sculpted by potentially quite general human learning mechanisms: patterns that afford optimal learnability are most likely to predominate in the languages of the world (e.g., Chomsky, 1965). In this case, predictive dependencies may inhabit phrases cross-linguistically precisely because they provide a useful cue to phrasal units. Explanations for the uniqueness of human language may reside, at least in part, in the uniqueness of human learning.

Acknowledgements

This research was funded by grants to J.S. from NICHD (R01HD37466) and NSF (BCS-9983630), by a core grant to the Waisman Center, University of Wisconsin-Madison from NICHD (P30 HD03352), and grants to MH from the McDonnell Foundation and NICHD (1 R01 DC005863-01A1). Thanks to R. Aslin, N. Chomsky, S. Curtin, T. Fitch, G. Marcus, M. MacDonald, S. Pollak, M. Seidenberg, and three anonymous reviewers for helpful comments on a previous draft.

References

- Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science*, *302*, 1234–1236.
- Braine, M. (1987). What is learned in acquiring words classes: A step toward acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Hillsdale, New Jersey: Erlbaum.
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*, 157–205.
- Christiansen, M. H., & Dale, R. (2004). The role of learning and development in the evolution of language. A connectionist perspective. In D. Kimbrough Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach. The Vienna Series in Theoretical Biology* (pp. 90–109). Cambridge, MA: MIT Press.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*, 377–380.
- Frigo, L., & McDonald, J. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, *39*, 218–245.
- Fragaszy, D. M., Galloway, A., Johnson-Pynn, J., & Brakke, K. (2002). The sources of skill in seriating cups in children, monkeys and apes. *Developmental Science*, *5*, 118–131.
- Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*.
- Gómez, R. L. (2006). Dynamically guided learning. In Y. Munakata & M. Johnson (Eds.), *Attention & Performance XXI: Processes of change in brain and cognitive development* (pp. 87–110). Oxford University Press.
- Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.
- Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language learning infants. *Psychological Science*, *17*, 670–674.
- Gómez, R. L., & LaKusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*, 567–580.
- Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Brain and Behavioral Sciences*, *14*, 531–595.
- Hauser, M. D., Barner, D., & O'Donnell, T. (2007). Evolutionary linguistics: A new look at an old landscape. *Language, Learning, and Development*, *3*, 101–132.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.
- Hauser, M., 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*, B41–B52.
- Hauser, M. D., Weiss, D. J., & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition*, *86*, B15–B22.
- Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77–80.
- McGonigle, B., Chalmers, M., & Dickinson, A. (2003). Nine item classification and seriation in cyberspace by *Cebus apella*. *Animal Cognition*, *6*, 185–197.
- Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, *30*, 678–686.
- Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*, 393–424.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*, 498–550.
- Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, *20*, 67–85.

- 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*. Somerville, MA: Cascadia Press.
- 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., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance: II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*, 85–117.
- Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*, 264–275.
- Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin and Review*, *12*, 307–313.
- Premack, D., & Premack, A. (2002). *Original intelligence*. New York: McGraw Hill.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology – General*, *118*, 219–235.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*, 493–515.
- Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*, 172–196.
- Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Savage-Rumbaugh, E. S. (1986). *Ape Language: From conditioned response to symbol*. New York: Columbia University Press.
- Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, *298*, 553–554.
- Smith, K. (1969). Learning co-occurrence restrictions: Rule learning or rote learning? *Journal of Verbal Behavior*, *8*, 319–321.