Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Psychology 49 (2004) 85–117
Cognitive Psychology
www.elsevier.com/locate/cogpsych

Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate

Elissa L. Newport,[a,*] Marc D. Hauser,[b] Geertrui Spaepen,[b] and Richard N. Aslin[a]

[a] *Department of Brain and Cognitive Sciences, University of Rochester, Meliora Hall, River Campus, Rochester, NY 14627, USA*
[b] *Department of Psychology and Program in Neurosciences, Harvard University, USA*

Accepted 1 December 2003
Available online 21 February 2004

Abstract

In earlier work we have shown that adults, infants, and cotton-top tamarin monkeys are capable of computing the probability with which syllables occur in particular orders in rapidly presented streams of human speech, and of using these probabilities to group adjacent syllables into word-like units. We have also investigated adults' learning of regularities among elements that are *not* adjacent, and have found strong selectivities in their ability to learn various kinds of non-adjacent regularities. In the present paper we investigate the learning of these same non-adjacent regularities in tamarin monkeys, using the same materials and familiarization methods. Three types of languages were constructed. In one, words were formed by statistical regularities between *non-adjacent syllables*. Words contained predictable relations between syllables 1 and 3; syllable 2 varied. In a second type of language, words were formed by statistical regularities between *non-adjacent segments*. Words contained predictable relations between consonants; the vowels varied. In a third type of language, also formed by regularities between *non-adjacent segments*, words contained predictable relations between vowels; the consonants varied. Tamarin monkeys were exposed to these languages in the same fashion as adults (21 min of exposure to a continuous speech stream) and were then tested in a playback paradigm measuring spontaneous looking (no reinforcement). Adult subjects learned the second and third types of language easily, but failed to learn the first. However, tamarin monkeys showed a different pattern, learning the first and third type of languages but not the

* Corresponding author.
E-mail address: newport@bcs.rochester.edu (E.L. Newport).

0010-0285/$ - see front matter
doi:10.1016/j.cogpsych.2003.12.002

second. These differences held up over multiple replications, using different sounds instantiating each of the patterns. These results suggest differences among learners in the elementary units perceived in speech (syllables, consonants, and vowels) and/or the distance over which such units can be related, and therefore differences among learners in the types of patterned regularities they can acquire. Such studies with tamarins open interesting questions about the perceptual and computational capacities of human learners that may be essential for language acquisition, and how they may differ from those of non-human primates.

1. Introduction

There are striking differences between human languages and animal communication systems—surprisingly, even between human and non-human primate vocalizations (Cheney & Seyfarth, 1990; Hauser, 1996; Hauser, Chomsky, & Fitch, 2002). One central difference concerns the combinatorial nature of human languages: all human languages, whether spoken or signed, and whether used in industrialized or agrarian societies, are built of small elements of form that combine, recursively and hierarchically, to make larger units; acoustic or manual features combine to form segments, which in turn combine to form syllables, words, phrases, and sentences (Chomsky, 1965, 1995). In order to learn such a combinatorial system, human children must be able to isolate some starting elements from the speech stream and then acquire the patterns of combination permitted in the particular language to which they are exposed. In contrast, it is not clear whether non-human primate vocalizations comprise a combinatorial system (Hailman & Ficken, 1987; Zuberbühler, 2002), or whether non-human primates are capable of such combinatorial learning (Fitch & Hauser, 2004; Greenfield, 1991; Hauser, 2000; Hauser, Newport, & Aslin, 2001; Savage-Rumbaugh et al., 1993). The present studies are part of a larger research program attempting to address such questions, by designing small learning problems modeled after parts of human languages and by asking whether human adults, human infants, and several species of non-human primates are capable of mastering this type of pattern learning from mere exposure. Our aim in these studies is therefore to see whether some of the differences that have permitted the development of human languages might have arisen from differences in the basic combinatorial and computational mechanisms humans bring to the task of language acquisition.

Our approach follows in the tradition of earlier research that has addressed similar questions at the level of perception. In particular, Kuhl and Miller (1975, 1978; see Kuhl, 1986, 1989, for an overview) conducted studies with chinchillas, using operant training techniques to ask whether animals with a basic mammalian auditory system much like that of humans would show the classic phenomena of human speech perception, such as categorical perception and context effects, and therefore whether these phenomena arose from general auditory mechanisms or were evolutionarily specialized for human speech. More recently, Kuhl (1991) has shown that non-human and human primates differ in the way speech categories are internally organized. Using spontaneous looking techniques to study similar problems, Hauser

and colleagues (Hauser et al., 2001, 2002; Ramus, Hauser, Miller, Morris, & Mehler, 2000) have revealed a number of similarities and differences across species.

In the present work, we ask not about perception, but rather about the abilities of non-human primates to perform some of the computational tasks involved in language learning—in particular, the learning of combinatorial patterns. Unlike some studies in the non-human primate literature, it is not our aim to ask whether animals can or cannot acquire a human language. Rather, we aim to investigate precisely where the perceptual and computational abilities required during learning may differ across species (and also across domains, such as language, music, and vision).

In recent work, we have shown that human adults, young children, and infants are capable of computing transitional probabilities[1] among adjacent syllables in rapidly presented streams of speech, and of using these statistics to group syllables into word-like units (Aslin, Saffran, & Newport, 1998; Saffran, Aslin & Newport, 1996; Saffran, Newport & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Such results have suggested that humans may engage in a set of computational processes that may be important not only for word segmentation, but perhaps as well for the acquisition of other aspects of human language (Newport & Aslin, 2004). We have therefore also been interested in asking whether non-human primates can engage in any of these processes, and where they might differ from human adults and infants in such abilities. In a previous series of studies (Hauser et al., 2001) we exposed cotton-top tamarins to the same continuous stream of synthetic human speech used in studies with human infants. We then asked whether, without explicit training or reinforcement, they too could acquire the statistics of which sounds followed each other in the stream and how frequently they did so. Like infants, they demonstrated this learning by showing a novelty response in looking to test stimuli: after exposure, tamarins were significantly less likely to look toward the playback speaker when they heard a word from the stream—a three-syllable sequence that occurred in the exposure stream with high probability—and more likely to look when they heard a partword—a three-syllable sequence that also occurred in the exposure stream, but with less consistency among the syllables.

This result revealed a number of basic abilities previously unknown in tamarins. In order to succeed on our learning task, tamarins had to discriminate, on some basis, 12 different human speech syllables; further, they had to track the temporal order in which these syllables occurred, and the frequency or probability of these orderings. These results, however, raise a number of questions about the tamarin perceptual system and its computational limits and open the door to a further comparative research enterprise. For example, what do tamarins perceive about human speech elements? Do they fully discriminate the 12 syllables from one another, or do they only

[1] More technically, we have shown that learners compute a conditionalized statistic which tracks the consistency with which elements occur together and in a particular order, baselined against individual element frequency. *Transitional probability* is a particular type of temporally ordered conditional probability, first used for psycholinguistic materials by Miller and Selfridge (1950). But our findings are also compatible with the claim that learners might be computing another closely related statistic, such as *mutual information* or *conditional entropy*.

perceive contrasts in the (louder and longer) vowels? Do they represent human speech in terms of holistic syllables (as has been argued for in young human infants; Bertoncini & Mehler, 1981; Jusczyk & Derrah, 1987; cf. Jusczyk, 1997, for discussion), or do they have a finer representation of speech, in terms of features and/or phonetic segments (as has been suggested for older infants and chinchillas; Hillenbrand, 1983, 1984; Kuhl & Miller, 1975; Kuhl, 1985, 1986)? In addition, there are important questions about the nature of the temporal order and statistical regularities tamarins are capable of perceiving and remembering. Although they readily extract regularities among immediately adjacent syllables, can they also extract regularities among non-adjacent syllables or segments? In short, where in these processes do humans and non-human primates begin to diverge?

In a series of recent studies (Newport & Aslin, 2004), we have built more complex languages than those used in our original research and have used them to ask these questions of human adults. In the present studies, we ask the same questions of cotton-top tamarins. The first study investigates whether tamarins can acquire an artificial language in which the words of the language are built out of regularities among non-adjacent syllables. The second and third studies examine whether tamarins can acquire artificial languages in which the words of the languages are built out of regularities among non-adjacent phonemic segments (in the second study, regularities among the consonants; in the third study, regularities among the vowels). One possibility is that tamarins will not be able to acquire any of these languages, but will be limited to learning only the simplest relations among adjacent units. If correct, these results would parallel ongoing pilot work with young human infants. On the other hand, our results with human adults show very selective learning of these different types of languages: some of these languages are easily learned, while others are not. An interesting possibility, therefore, is that tamarins may show their own selectivities of learning: perhaps like those of humans, but perhaps different, reflecting (as do the adult human results) that only particular types of elements are perceived and that only particular types of patterned relations can be computed and learned.

Human adults are not readily able to acquire the first type of language, whose regularities occur among non-adjacent syllables (Newport & Aslin, 2004). This is despite the fact that these languages are very similar to our original languages in the sounds they utilize and many aspects of the statistics across the stream; the only difference from our original studies is in the non-adjacency between related syllables. Importantly, natural human languages of the world also do not commonly form words from a stem consisting of related syllables 1 and 3, and other syllables inserted in the middle (see Newport & Aslin, 2004, for discussion). In contrast, human adults are readily able to acquire both the second and third types of languages, whose regularities occur among non-adjacent phonemic segments (either consonants or vowels). Consistent with this finding, many natural human languages display patterns of this type (for example, consonant patterns in Hebrew and Arabic; vowel harmony patterns in Turkish). An interesting possibility is thus that human languages have been shaped by the same selectivities of learning shown in our experiments.

What types of patterns do tamarins display in their own natural vocalizations? A number of studies have suggested that tamarin vocalizations contrast noisy with

open vocal tract sounds and may be comprised of a number of syllable-like units in particular temporal sequences (Cleveland & Snowdon, 1981). But these studies are still in the early stages of understanding how tamarins themselves perceive their vocalizations and what types of patterns are discriminated (cf. Ghazanfar, Smith-Rohrberg, Pollen, & Hauser, 2002; Miller, Miller, Gil-da-Costa, & Hauser, 2001; Weiss, Garibaldi, & Hauser, 2001; Weiss & Hauser, 2002). At the same time, we do know that tamarins display surprising abilities to discriminate the number of elements in both auditory and visual stimuli, to retain elements of such stimuli in working memory, and to acquire simple abstract patterns among human speech sounds (Hauser, Dehaene, Dehaene-Lambertz, & Patalano, 2002; Hauser, Weiss, & Marcus, 2002). We therefore started the present studies of tamarin abilities, utilizing human speech materials, without knowing what precise pattern of results to expect, but aimed at utilizing a structured and fairly complex set of stimulus materials to further investigate statistical learning in a non-human species.

2. Experiment 1: Non-adjacent syllables

In the languages with adjacent regularities that were previously used with tamarins (Hauser et al., 2001; Saffran, Aslin & Newport, 1996), four tri-syllabic words were built from 12 different syllables, each used in only one word, with words following each other at random (excluding immediate repeats). In this design, the transitional probabilities between syllables within a word were 1.0; the transitional probabilities at word boundaries were .33. This type of language can be learned by 8-month-old human infants and adults (Saffran, Aslin & Newport, 1996; Saffran, Newport & Aslin, 1996), and by adult tamarin monkeys (Hauser et al., 2001). Experiment 1 uses the same logic as our earlier experiments, mirroring both the design and testing apparatus, but asks instead about computations over non-adjacent syllables. In order to build a language using non-adjacent syllable regularities, but with approximately the same level of structural complexity as in our studies of adjacent regularities, we created a language in which there were three regular word frames—sets of syllable pairs that co-occurred with a transitional probability of 1.0. However, we made these syllable pairs non-adjacent, inserting one of two different syllables in the middle. The same two middle syllables could occur in the middle of all three non-adjacent word frames. The words in this language thus all followed a 1-X-3 pattern, with three sets of 1–3 instantiations and two Xs, for a total of six different words in the language. The top portion of Table 1 presents a schematic of this word structure.

Given a stream of words, randomly ordered (excluding immediate repeats), following this pattern, there is no strong grouping of syllables if only adjacent syllable relations are computed: the adjacent transitional probabilities in such a stream vary from .5 (at the transition from syllable 1 to syllable X, and also at the transition from syllable 3 to syllable 1 of the next word) to .33 (at the transition from syllable X to syllable 3), with no extremely high or extremely low transitions at any point. However, there is a strong grouping of syllables if learners are able to compute non-adjacent syllable relations: among syllables 1 and 3 (one syllable away from another),

Table 1
Design of two non-adjacent syllable languages used in Experiment 1

CV_1	$[CV_2]$ $[CV_3]$	CV_4
CV_5	$[CV_2]$ $[CV_3]$	CV_6
CV_7	$[CV_2]$ $[CV_3]$	CV_8

1st–3rd Syllable word-frames	*2nd Syllables*
Language A	
di__tae	ki
po__ga	gu
ke__bu	
Language B	
bae__ku	pa
te__da	be
go__pi	

the transitional probability is 1.0, whereas the transitional probability among other syllables one away (syllable X and syllable 1 of the next word, or syllable 3 and syllable X of the next word) is only .33 or .5. Exposing subjects to such a stream can thus permit us to ask whether learners are readily able to compute such non-adjacent syllable statistics.

In order to be sure we are asking whether learners can acquire the type of structure we are investigating, and not merely responding on the basis of preferences or perceptual grouping among a specific set of sounds, this pattern was built into two different instantiations, called Language A and Language B. These Language instantiations were exactly the same in statistical structure and differed only in the assignment of particular phonetic elements to positions in the words of the languages. Half the tamarins were exposed to Language A, and half were exposed to Language B.

2.1. Method

2.1.1. Subjects

Subjects were 14 adult cotton-top tamarins (*Saguinus oedipus*), eight females and six males. This species is native to the rainforests of Colombia. All subjects were born in captivity at the New England Regional Primate Research Center, Southborough, MA or the Primate Cognitive Neuroscience Lab, Harvard University; they have been housed at Harvard since 1992. Animals live in social groups consisting of a mated pair, and in some cases, their offspring. Tamarins' frequency sensitivity includes the range of human speech sounds (Cleveland & Snowdon, 1981; Stebbins, 1983).

All subjects have experience in playback experiments involving both species-typical vocalizations and speech (Ramus et al., 2000; Weiss et al., 2001), and in experiments involving other cognitive and perceptual abilities (Hauser, 1997, 1998; Santos & Hauser, 1999). Because of this experience, subjects voluntarily move in

and out of their home cage and into a test area. When they arrive in the testing area, they are calm and will typically remain so for approximately 30 min. We can therefore present stimuli over a relatively long period of time without distressing them. All 14 subjects were tested in one of the two conditions, and all provided usable data.

2.1.2. Stimulus materials

Stimuli consisted of the same streams of synthetic speech-syllables used in the human adult studies of Newport and Aslin (2004), but synthesized at a 10% slower rate of speech for use with human infants and tamarins.[2] Table 1 shows the structure of the two languages. One speech stream (Language A) consisted of a 21-min random ordering of six three-syllable nonsense words (dikitae, digutae, pokiga, poguga, kekibu, and kegubu).[3] A second speech stream (Language B) consisted of a similarly structured stream of six different words (baepaku, baebeku, tepada, tebeda, gopapi, gobepi). To form this 21-min stream, eight blocks, each consisting of a random ordering of two tokens of each of the six words, were concatenated into a text in random order, with the stipulation that the same word, and the same word-frame, never occurred twice in a row. All word boundaries were removed from the text, rendering a list of 288 syllables. The text was then read by the MacInTalk speech synthesizer, using the text-to-speech application Speaker, running on a Power Macintosh G3 computer. Because the synthesizer was not informed of word boundaries, it did not produce any acoustic word boundary cues and produced equivalent levels of coarticulation between all syllables. The speech stream contained no pauses and was produced by a synthetic female voice (Victoria) in monotone. The output of the synthesizer was recorded to Sony minidisc directly from the sound output of the Power Macintosh computer and then recorded again into SoundEdit 16 to obtain a digital audio file. Once recorded in SoundEdit, each syllable was edited to .22–.25 s in length. The speech stream contained no pauses and played at a rate of 4.46 and 4.20 syllables per second, for Language A and Language B, respectively (268 and 252 syllables per minute). This 1-min stream of speech was looped to form the 21-min exposure stream on day 1 and the 2-min re-exposure on day 2.

The tamarins were exposed to the 21-min stream in a room that was physically and acoustically isolated from their home room. All subjects sat in a cage and were passively exposed. We used an Advent powered partner speaker to broadcast the stream from a Macintosh Powerbook, at 70–75 dB SPL as measured 1 m away (6–10 dB louder than what is used for playback experiments involving their own vocalizations). The 2-min re-exposure was conducted in a sound attenuating chamber, with the stream broadcast from an Alesis speaker at the same loudness; during re-exposure, subjects sat in the test chamber and were fed small pieces of Froot Loops.

For both streams, then, there were no acoustic cues at word boundaries. The only available information for extracting words was the greater statistical regularity of

[2] The stimuli for this and all subsequent experiments were also tested with human adults and gave the same results as those with the faster adult stimuli reported in Newport and Aslin (2004); see Appendix A.

[3] Notation in the tables and text for describing the sounds used in our materials is in the International Phonetic Alphabet (IPA).

non-adjacent syllable sequences within words than of syllable sequences that spanned a word boundary. The learning of this statistical coherence of the non-adjacent syllables within words was tested by asking whether subjects could discriminate the *words* from *partwords* (three-syllable sequences that also occurred in the stream but spanned a word boundary). For each language, there were two test words and two test partwords. For Language A, the test words were digutae and pokiga; the test partwords were bupoki and taekegu. For Language B, the test words were baepaku and gobepi; the test partwords were dagobe and kutepa.[4] Test words and partwords were synthesized and edited in the same way as described for the streams above, except that both were generated by having MacInTalk produce these three-syllable items in isolation. This produced a falling intonation on the final syllable of each item, making each test item (words and partwords) sound like a word spoken in isolation.

Within the 21-min stream of speech, each of the six trisyllabic nonsense words occurred equally often and in random order, with the constraint that no word, and no word-frame, was immediately repeated. The transitional probability for non-adjacent syllables inside words (between syllables 1 and 3) was therefore 1.0; the transitional probability for non-adjacent syllables within part-words was .5 (between syllable 3 of one word and syllable 2 of the next). In order to discriminate between the test items, subjects would have to compute the transitional probability between non-adjacent syllables, or another closely related statistic.

Test items did not differ in the frequency of individual syllables. Importantly, they also did not differ in the transitional probabilities among adjacent syllables. The languages were designed so that they could not be learned by computing transitional probabilities only among adjacent syllables: as already noted, the adjacent transitional probabilities within words are .5 (from syllable 1 to syllable X) and .33 (from syllable X to syllable 3), while those within partwords are .5 (from syllable 3 to syllable 1) and .5 (from syllable 1 to syllable X). These adjacent transitional probabilities thus have no extremely high or extremely low transitions at any point, and the differences between the two types of items would not be ones we would expect learners to discriminate.[5]

2.1.3. Procedure

Each of two sets of sessions involved the same 2-day familiarization-test procedure, and differed only in the stimuli presented for familiarization and test. Half the colony

[4] In order to balance consonants and vowels used within the four test items (as compared with the larger number of test items used with human adults), the test items used in this and subsequent experiments were not precisely the same as the ones used in Newport and Aslin (2004) with adults. However, as noted above, for this and all subsequent experiments the stimuli were also tested with human adults and gave the same results as those reported in Newport and Aslin (2004); see Appendix A.

[5] However, if subjects are able to perform a discrimination on the basis of the .5 versus .33 contrast, they should favor the consistency of partwords over the words; such an effect would appear as a high likelihood of looking in response to the words (since previous results in this paradigm always show a novelty effect). In contrast, the non-adjacent transitional probabilities of 1.0 within words versus .5 within partwords should lead tamarins to favor the words over the partwords, an effect which should appear as a high likelihood of looking in response to the partwords.

of tamarins was exposed to and tested on Language A, while the other half was exposed to and tested on Language B. For each language, on day 1, the relevant tamarins were familiarized to a 21-min continuous speech stream. On day 2, each of these tamarins was placed individually in the sound proof chamber (see Hauser et al., 2001, for an illustration) and re-familiarized with a 2-min corpus of the speech stream, followed immediately by eight test trials: one instance of each of the two words and the two partwords in random order, followed by another instance of each of the two words and the two partwords in a different random order. Inter-trial intervals ranged from 10 to 60 s. We did not run subjects who failed to leave their home room cage on the day of testing, or those who jumped around the test cage and failed to sit quietly.

No explicit training or reinforcement was provided during any phase of the experiment, during either familiarization or test. The dependent measure during testing was an orienting response to a stimulus presented from a concealed loudspeaker. For clarity in scoring a response, stimuli were presented when the subject was still and faced 180° away from the concealed speaker (that is, looking down and away from the speaker; see Hauser et al., 2001). Subjects were scored as responding if they turned and looked in the direction of the speaker during the presentation of the test stimulus and continued to look after it ended (to ensure that they were responding to the full stimulus and not just to its onset), or if they turned and looked in the direction of the speaker within 2 s after the test stimulus ended. Orienting responses were scored independently by two observers from digitized video recordings, blind to the test condition; in all of the experiments reported in this paper, reliabilities ranged between 0.90 and 0.98. Trials on which subjects were not facing away from the speaker at stimulus onset, or on which observers could not code a response unambiguously, were eliminated (15%). This procedure has been used reliably in other experiments (Hauser et al., 2001; Ramus et al., 2000).

2.2. Results

Each animal's responses in each test condition (words vs. partwords) were converted to a percentage, and these percentages were averaged across animals. Fig. 1 presents the mean percent of trials showing an orienting response to words vs. partwords. Both words and partwords occurred in the familiarization stream; but only the words included the high-probability non-adjacent syllable relations. Comparing subjects' responses to words vs. partwords thus assesses their learning of this non-adjacent syllable probability. Tamarins were significantly more likely to orient to partwords than to words (Wilcoxon Signed Ranks, $z(8) = -2.31$, $p = .02$ for Language A, $z(6) = -1.753$, $p = .08$ for Language B, $z(14) = -2.87$, $p = .004$ for A and B combined across each animal's useable test trials).

2.3. Discussion

These results show that tamarins are able to acquire dependencies between non-adjacent syllables, just as they had been able to acquire dependencies between

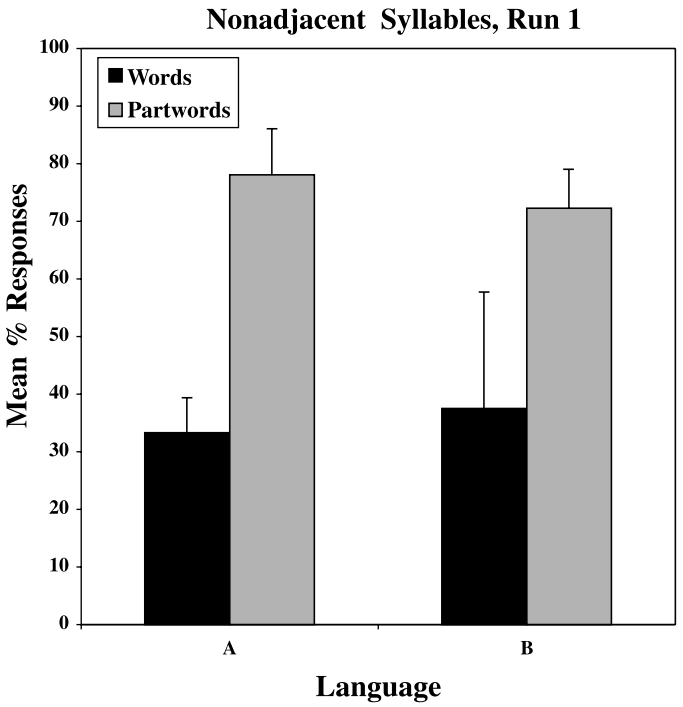

Fig. 1. Mean percent orienting responses to words versus partwords, for non-adjacent syllable languages (Language A on left, Language B on right). Data are from Experiment 1 (Run 1).

adjacent syllables in our earlier results (Hauser et al., 2001). The present results thus provide one more piece of evidence, consistent with our prior findings, that tamarins are able to process fairly complex statistical aspects of human speech streams. The findings of the present study go beyond our prior results in showing that they can keep track of non-adjacent elements as well as adjacent ones.

One very surprising aspect of these findings, however, is that human adults are not able to succeed on this task. While humans can perform other non-adjacent element computations (see below), they apparently find computations of non-adjacent syllables extremely difficult. Before attempting to make sense of this difference between tamarins and humans, it seemed important to be sure this surprising result could be replicated. We therefore ran the same experiment again, 7 months later, without exposing the animals to the materials in the intervening time.

3. Experiment 1A: Second run, non-adjacent syllables

This experiment was identical to Experiment 1 and was conducted to see whether those results were reliable.

3.1. Method

3.1.1. Subjects

The same 14 tamarins participated in this experiment as in Experiment 1, but divided differently into two groups for exposure to the two language instantiations. A total of six (three males, three females) provided usable data for Language A, and eight (four males, four females) for Language B.

The apparatus, stimuli, and procedure were the same as in Experiment 1.

3.2. Results

As before, each animal's responses in each test condition (words vs. partwords) were converted to a percentage, and these percentages were averaged across animals. Fig. 2 presents the mean percent of trials showing an orienting response to words vs. partwords. Tamarins were significantly more likely to orient to partwords than to words (Wilcoxon Signed Ranks, $z(6) = -2.20$, $p = .028$ for Language A, $z(8) = -2.38$, $p = .017$ for Language B, $z(14) = -3.23$, $p = .0012$ for A and B combined across each animal's useable test trials).

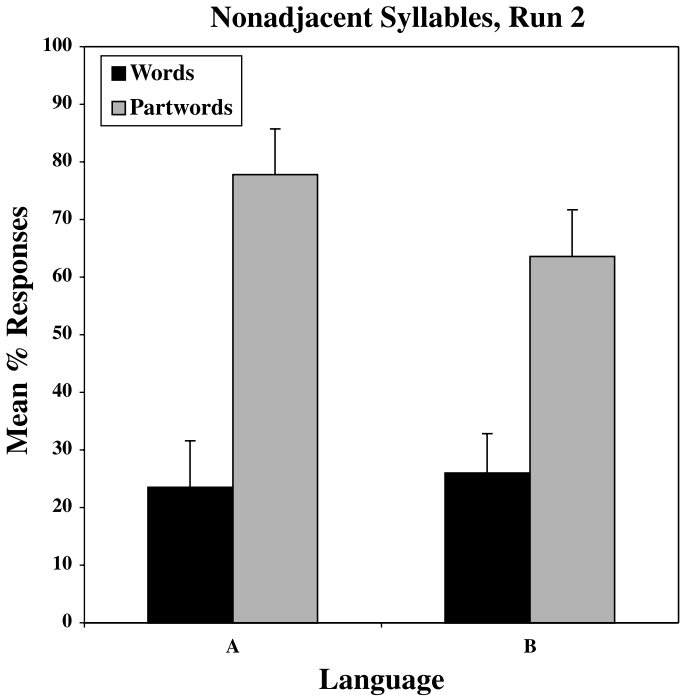

Fig. 2. Mean percent orienting responses to words versus partwords, for non-adjacent syllable languages (Language A on left, Language B on right). Data are from Experiment 1A (Run 2).

3.3. Discussion

These results replicate precisely the results obtained in Experiment 1. Altogether our subjects have performed a successful discrimination on four different runs: two different languages A and B, each exhibiting the same non-adjacent syllable dependencies, each run twice on different subgroups of the tamarin population. It is apparently the case, then, that tamarins can compute statistical patterns among non-adjacent syllables in a stream of human speech and can use these patterns to group syllables into trisyllabic units ('words').

As noted above, human adults are not able to perform this same task, using the same materials, exposed for the same length of time, and tested in similar ways. Newport and Aslin (2004) ran this experiment with adults, exposing them for 21 min to a stream with the same structure, and then asked them to perform a two-alternative forced choice task, choosing whether words or partwords formed a better group or unit. Adults were not able to perform this discrimination. Why are tamarins able to perform this particular non-adjacent computation when human adults find the same computation, using the same materials and a very similar task, extremely difficult?

Given the other cognitive differences between tamarins and humans, it seems extremely unlikely that this difference in outcomes is due to greater information handling capabilities in tamarins than in human adults: for example, a greater ability to bridge distance across the speech stream or a greater short-term memory span. More likely, the difference must be due to the fact that the speech signal is represented or processed in a different way by tamarins than by humans. To explore this difference further, we proceeded to test tamarins on two other types of languages involving non-adjacent dependencies, both of which we had studied with humans and on both of which humans succeeded in demonstrating non-adjacent computational abilities. Both of these sets of materials involve *non-adjacent phonemic segments*, rather than *non-adjacent syllables*. In Experiment 2, we test tamarins on their ability to compute non-adjacent segment regularities involving consonants, and in Experiment 3 we test tamarins on their ability to compute non-adjacent segment regularities involving vowels.

4. Experiment 2: Non-adjacent phonemic segments (consonants)

In our experiments with human adults (Newport & Aslin, 2004) we noted that human languages do not frequently exhibit word-formation patterns made of non-adjacent syllables, like those we tested in Experiment 1.[6] However, a common non-adjacent pattern in human languages involves regular relationships among

[6] Natural languages do not commonly construct words out of a non-adjacent syllable pattern; most phonological patterns occur among adjacent phonetic segments or syllables. In some languages, such as Tagalog, words may contain two-syllable stems, with certain inflections that can be inserted between the two syllables. But the two syllables also often occur adjacent to one another. In such languages, the adjacent pattern (or production of just one of the syllables) occurs early in acquisition, while acquisition of the non-adjacent pattern is much later (Slobin, 1973) and likely occurs via the earlier control over the adjacent pattern. See Newport and Aslin (2004) for further discussion.

non-adjacent phonemic segments. For example, Semitic languages like Hebrew and Arabic form many words out of a three-consonant stem, such as *k-t-b* ('to write'). The vowels inserted between these consonants then vary, depending on whether the word is present vs. past tense, or active vs. passive. In order to acquire words of this type, learners would have to keep track of the consistent pattern among the consonants, ignoring the variations among the vowels. In a second series of experiments in Newport and Aslin (2004) we synthesized materials exhibiting this type of pattern, to ask whether human learners were capable of acquiring this particular non-adjacent regularity. In contrast to our previous results on non-adjacent syllable regularities, human learners (adult English speakers) were readily able to acquire this non-adjacent segment regularity. In the present experiment we ask whether tamarins can do the same.

In order to test this type of structure in as simple a way as possible, we formed two three-consonant frames, with two different vowels possible in each of the vocalic positions. Table 2 shows an illustration of this word structure. Given this type of structure, the transitional probabilities between the consonants within a word are 1.0; the transitional probabilities between the consonants across word boundaries are .5. However, the word structure and the stream ordering rules were carefully designed so that no adjacent transitional probability computation, either between adjacent syllables or between adjacent segments, would produce a coherent grouping. The transitional probabilities between adjacent *syllables* within words are .5. In order to make the transitional probabilities between adjacent syllables across word boundaries also equal to .5 (so that words could not be learned by computing these adjacent syllable statistics), the speech streams for these languages were created following the rule that a particular word-final syllable would always be followed by one of two (of the possible four) word-initial syllables. Words were also never immediately repeated. The transitional probabilities between adjacent segments (from consonant to vowel and vowel to consonant) were also .5 all along the stream, with no high or low adjacent transitions that would permit learners to form words or groups by these computational methods. In short, then, given a stream of words following these patterns, there is no grouping of syllables into words if adjacent syllable or adjacent

Table 2
Design of two non-adjacent segment (consonant) languages used in Experiment 2

$C_1[v_1]$ $C_2[v_3]$ $C_3[v_5]$
$[v_2]$ $[v_4]$ $[v_6]$

$C_4[v_1]$ $C_5[v_3]$ $C_6[v_5]$
$[v_2]$ $[v_4]$ $[v_6]$

Consonant-frames	Vowel-fillers
Language A	
p_g_t_	[_a] [_i] [_ae]
d_k_b_	[_o] [_u] [_e]
Language B	
t_d_k_	[_ae] [_a] [_i]
b_p_g_	[_e] [_o] [_u]

segment relations are computed. However, words can readily be learned if non-adjacent (consonant) segment regularities are computed.

While this pattern forms 16 different words in the language, the inventory of sounds used and the size of the transitional probability differences to be acquired were very similar to those of the languages used in Experiment 1. (See Newport & Aslin, 2004, for further discussion of the details of these languages.)

As in Experiment 1, in order to be sure we are asking whether learners can acquire the type of structure we are investigating and are not merely responding on the basis of preferences or perceptual grouping among a specific set of sounds, we created two different versions of this pattern, called Language A and Language B. These two versions were exactly the same in statistical structure and differed only in the assignment of particular phonetic elements to positions in the words of the languages. Half the tamarins were exposed to Language A, and half were exposed to Language B.

4.1. Method

4.1.1. Subjects

The same 14 tamarins participated in this experiment as in our previous experiments, but divided differently into two groups for exposure to the two language instantiations. We also added two more animals, previously untested. Three tamarins were eliminated from the experiment due to poor health or poor behavior during testing; a total of seven (three males, four females) therefore provided usable data for Language A, and six (three males, three females) for Language B.

4.1.2. Stimulus materials

Stimuli consisted of the same streams of synthetic speech-syllables used in the human adult studies of Newport and Aslin (2004), but synthesized at a 10% slower rate of speech for use with human infants and tamarins. (See Appendix A for evidence that this change of rate does not alter the results for human adults.) As shown in Table 2, one speech stream (Language A) consisted of a 21-min constrained random ordering of 16 three-syllable nonsense words. A second speech stream (Language B) consisted of a similarly structured stream of sixteen different words. To form this 21-min stream, six blocks, each consisting of a different random ordering of the 16 words, were concatenated into a text in a constrained random order, with the stipulation that the same word never occurred twice in a row and each word-final syllable could only be followed by either of two particular word-initial syllables. All word boundaries were removed from the text, rendering a list of 288 syllables. The text was then read by the MacInTalk speech synthesizer, using the text-to-speech application Speaker, running on a Power Macintosh G3 computer, with all synthesis, re-recording, and editing done as in Experiment 1. The speech stream contained no pauses and played at a rate of 4.26 and 4.22 syllables per second, for Language A and Language B, respectively (255 and 253 syllables per minute). This 1-min stream of speech was looped to form the 21-min exposure stream on day 1 and the 2-min re-exposure on day 2.

For both streams, then, there were no acoustic cues at word boundaries. The only available information for extracting words was the greater statistical regularity of

non-adjacent segment sequences within words than of segment sequences that spanned a word boundary. The learning of this statistical coherence of the non-adjacent segments within words was tested by asking whether subjects could discriminate the *words* from *partwords* (three-syllable sequences that also occurred in the stream but spanned a word boundary); as for Experiment 1, we expected subjects to look more often at partwords than words if they extracted the appropriate statistics. For each language, there were two test words and two test partwords. For Language A, the test words were dokibae and pagute; the test partwords were bepogi and taedaku. For Language B, the test words were bepogi and taedaku; the test partwords were gutedo and kibaepa. As in Experiment 1, test words and partwords were synthesized and edited in the same way as described for the streams above, except that both were generated by having MacInTalk produce these three-syllable items in isolation. This produced a falling intonation on the final syllable of each item, making each test item (words and partwords) sound like a word spoken in isolation.

Within the 21-min stream of speech, each of the sixteen trisyllabic nonsense words occurred equally often, in a constrained order such that no word was immediately repeated and each word-final syllable could only be followed by either of two word-initial syllables. The transitional probabilities for non-adjacent segments inside words (between the consonants: segments 1, 3, and 5) were therefore 1.0; the transitional probabilities for non-adjacent segments within part-words were .5 (between segment 5 of one word and segment 1 of the next) and 1.0 (between segments 1 and 3). In order to discriminate between the test items, subjects would have to compute the transitional probability between non-adjacent segments, or another closely related statistic.

Test items did not differ in the frequency of individual segments or syllables. Importantly, they also did not differ in the transitional probabilities among adjacent segments or syllables. The languages were designed so that they could not be learned by computing transitional probabilities only among adjacent elements: as already noted, the adjacent transitional probabilities within and between words are all .5.

The apparatus and procedures were the same as in Experiment 1.

4.2. Results

Each animal's responses in each test condition (words vs. partwords) were converted to a percentage, and these percentages were averaged across animals. Fig. 3 presents the mean percent of trials showing an orienting response to words vs. partwords. Tamarins did not show any tendency to orient more to partwords than to words, on either Language A or Language B (Wilcoxon Signed Ranks, $z(7) = -.105$, $p = .92$, ns, for Language A, $z(6) = -.539$, $p = .59$, ns, for Language B, $z(13) = -.40$, $p = .69$, ns, for A and B combined across each animal's useable test trials).

4.3. Discussion

While tamarins were able to compute patterned relations between non-adjacent syllables in Experiment 1, they apparently do not perform such a computation between non-adjacent segments (consonants). Again, however, before interpreting

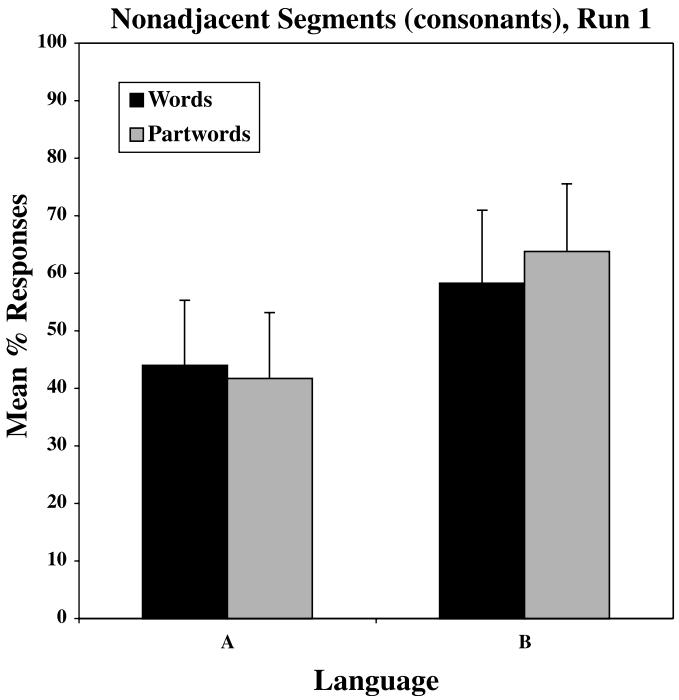

Fig. 3. Mean percent orienting responses to words versus partwords, for non-adjacent segment (consonant) languages (Language A on left, Language B on right). Data are from Experiment 2 (Run 1).

these results more fully, we decided to run the experiment a second time, to determine whether the results were reliable. We therefore ran the same experiment, approximately 2 months later, without exposing the animals to the materials in the intervening time.

5. Experiment 2A: Second run, non-adjacent segments (consonants)

This experiment was identical to Experiment 2 and was conducted to see whether those results were reliable.

5.1. Method

5.1.1. Subjects

The same 16 tamarins participated in this experiment as in Experiment 2, plus an additional two naïve animals, but divided differently into two groups for exposure to the two language instantiations. Three tamarins were excluded due to either poor health or poor behavior during the test session. A total of eight subjects (four females, four males) provided usable data for Language A, and 7 (three males, four females) for Language B.

5.1.2. Stimulus materials

The familiarization materials were the same as in Experiment 2. Test items were changed slightly to see whether a different selection of the test words and partwords would produce a better result. For Language A, the test words were dokibae and pagute; the test partwords were bepogi and taedaku. For Language B, the test words were baepagu and tedoki; the test partwords were gitaeda and kubepo.

The apparatus and procedures were the same as in Experiment 2.

5.2. Results

Each animal's responses in each test condition (words vs. partwords) were converted to a percentage, and these percentages were averaged across animals. Fig. 4 presents the mean percent of trials showing an orienting response to words vs. partwords. Tamarins did not show any tendency to orient more to partwords than to words (Wilcoxon Signed Ranks, $z(8) = -.73$, $p = .4652$, ns, for Language A, $z(7) = -.809$, $p = .4185$, ns, for Language B, $z(15) = -.296$, $p = .7671$, ns, for A and B combined across each animal's useable test trials).

5.3. Discussion

These results precisely replicate those obtained in Experiment 2. Across a total of four runs on our non-adjacent segment (consonants) materials, then, it appears that the tamarins cannot learn this type of pattern.[7] Importantly, these are the same subjects who successfully discriminated the non-adjacent syllables in Experiment 1.

Across Experiments 1 and 2, we have found that tamarins can acquire patterns among non-adjacent syllables, but not among non-adjacent segments, at least when the patterned segments are consonants. One possible reason for the tamarins' failure on consonant patterns is that perhaps they cannot perceive any sub-syllabic elements and hear human speech only in terms of larger and more holistic chunks. Another problem might be with stop consonants in particular. Stop consonants are known to be especially difficult for human adults, due to the brevity of their acoustic cues and the variability of these cues across differing vowel contexts (Liberman, 1970; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Furthermore, non-human primates, and non-human animals more generally, do not naturally produce consonant-like sounds in their vocal repertoires, due in part to the structure of the supralaryngeal vocal tract and the flexibility of their articulators (Hauser, 1996; Lieberman, 1984). Alternatively, it is possible that, because we have repeatedly tested

[7] Between Runs 1 and 2 of Experiment 2, we also ran another version of Experiment 2, using the same streams as in the other runs, but testing tamarins on an easier contrast involving words versus non-words (rather than partwords). Nonwords consist of three syllables from the language in an order that never occurred in the familiarization corpus. In our studies with humans, the word/nonword contrast is easier and sometimes reveals learning when the more difficult word/partword contrast does not. However, the results of this run with tamarins were the same as those from the word/partword contrast: no significant learning of non-adjacent segment (consonant) regularities.

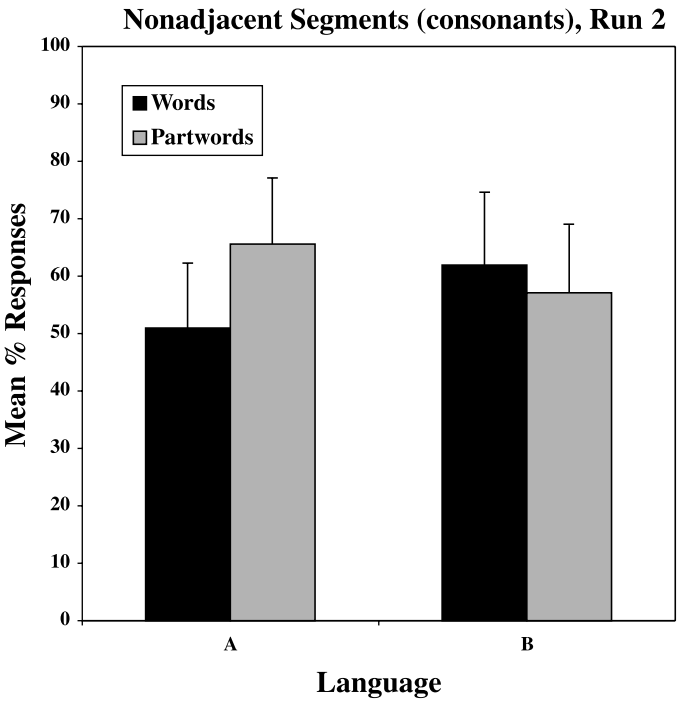

Fig. 4. Mean percent orienting responses to words versus partwords, for non-adjacent segment (consonant) languages (Language A on left, Language B on right). Data are from Experiment 2A (Run 2).

the same subjects on similar types of materials, their overall level of responding and attentiveness has decreased. Although the overall levels of responding in Experiments 1 and 2 were comparable, it is important to assess whether these subjects can now, following a failure to discriminate, show successful discrimination with another set of stimuli. In our final experiment, then, we tested tamarins on a non-adjacent segment pattern involving vowels rather than consonants.

6. Experiment 3: Non-adjacent phonemic segments (vowels)

We noted above that human languages frequently exhibit word-formation patterns made of non-adjacent phonemic segments. In Experiment 2, we tested tamarins on a non-adjacent segment pattern involving consonants, like that of Hebrew or Arabic. However, another common non-adjacent pattern in human languages involves regular relationships among vowels. For example, in Turkish, the vowels across a word must agree with one another in certain features, such as place of articulation or roundness. (This process is called 'vowel harmony.') In acquiring such a language, learners would have to keep track of the consistent pattern among the vowels, ignoring variations in the intervening consonants. In a third series of experiments in Newport and Aslin (2004) we synthesized materials exhibiting this type of pattern, and

found that human learners (adult English speakers) were indeed able to acquire this type of non-adjacent segment regularity. But so far our results with tamarins are extremely different from those with humans: tamarins can acquire non-adjacent syllable regularities, while humans do not; tamarins cannot acquire non-adjacent segment regularities involving consonants, while humans do. In the present experiment we ask whether tamarins can acquire non-adjacent segment regularities involving vowels.

The structure of this type of language is identical to that used for non-adjacent segments involving consonants, except that in this case the vowels are predictable while the consonants vary (rather than the reverse). In order to test this type of structure in as simple a way as possible, we formed two three-vowel frames, with two different consonants possible in each of the consonantal positions. Table 3 shows an illustration of this word structure. Given this type of structure, the transitional probabilities between the vowels within a word are 1.0; the transitional probabilities between the vowels across word boundaries are .5. However, the word structure and the stream ordering rules were carefully designed so that no adjacent transitional probability computation, either between adjacent syllables or between adjacent segments, would produce a coherent grouping. The transitional probabilities between adjacent *syllables* within words are .5. In order to make the transitional probabilities between adjacent syllables across word boundaries also equal to .5 (so that words could not be learned by computing these adjacent syllable statistics), the speech streams for these languages were created following the rule that a particular word-final syllable would always be followed by either of two (of the possible four) word-initial syllables. Words were also never immediately repeated. The transitional probabilities between adjacent segments (from consonant to vowel and vowel to consonant) were also .5 all along the stream, with no high or low adjacent transitions that would permit learners to form words or groups by these computational methods. In short, then, given a stream of words following these patterns, there is no grouping of syllables into words if adjacent syllable or adjacent segment relations are computed. However, words can readily be learned if non-adjacent (vowel) segment regularities are computed.

Table 3
Design of two non-adjacent segment (vowel) languages used in Experiment 3

$[c_1]V_1$ $[c_3]V_2$ $[c_5]V_3$
$[c_2]$ $[c_4]$ $[c_6]$

$[c_1]V_4$ $[c_3]V_5$ $[c_5]V_6$
$[c_2]$ $[c_4]$ $[c_6]$

Vowel-frames	Consonant-fillers
Language A	
_a_u_e	[p_] [g_] [t_]
_o_i_ae	[d_] [k_] [b_]
Language B	
_ae_a_u	[t_] [d_] [k_]
_e_o_i_	[b_] [p_] [g_]

While this pattern forms 16 different words in the language, the inventory of sounds used and the size of the transitional probability differences to be acquired are very similar to those of the languages used in Experiment 1 and exactly the same as those of the languages used in Experiment 2. (See Newport & Aslin, 2004, for further discussion of the details of these languages.)

As in Experiments 1 and 2, in order to be sure we are asking whether learners can acquire the type of structure we are investigating, and not merely responding on the basis of preferences or perceptual grouping among a specific set of sounds, this pattern was built in two different instantiations, called Language A and Language B. These Language instantiations were exactly the same in statistical structure and differed only in the assignment of particular phonetic elements to positions in the words of the languages. Half the tamarins were exposed to Language A, and half were exposed to Language B.

6.1. Method

6.1.1. Subjects

The same 18 tamarins tested in Experiment 2 participated in this experiment, plus one additional naïve animal, but divided differently into two groups for exposure to the two language instantiations. One tamarin was eliminated due to poor behavior during testing; a total of nine (three males, six females) therefore provided usable data for Language A, and nine (five males, four females) for Language B.

6.1.2. Stimulus materials

Stimuli consisted of the same streams of synthetic speech-syllables used in the human adult studies of Newport and Aslin (2004), but synthesized at a 10% slower rate of speech for use with human infants and tamarins. (See Appendix A for evidence that this change of rate does not alter the results for human adults.) As shown in Table 3, one speech stream (Language A) consisted of a 21-min constrained random ordering of sixteen three-syllable nonsense words. A second speech stream (Language B) consisted of a similarly structured stream of sixteen different words. To form this 21-min stream, six blocks, each consisting of a different random ordering of the 16 words, were concatenated into a text in a constrained random order, with the stipulation that the same word never occurred twice in a row and each word-final syllable could only be followed by either of two particular word-initial syllables. All word boundaries were removed from the text, rendering a list of 288 syllables. The text was then read by the MacInTalk speech synthesizer, using the text-to-speech application Speaker, running on a Power Macintosh G3 computer, with all synthesis, re-recording, and editing done as in Experiment 1. The speech stream contained no pauses and played at a rate of 4.31 and 4.27 syllables per second, for Language A and Language B, respectively (259 and 256 syllables per minute). This 1-min stream of speech was looped to form the 21-min exposure stream on day 1 and the 2-min re-exposure on day 2.

For both streams, then, there were no acoustic cues at word boundaries. The only available information for extracting words was the greater statistical regularity of non-adjacent segment sequences within words than of segment sequences that spanned

a word boundary. The learning of this statistical coherence of the non-adjacent segments within words was tested by asking whether subjects could discriminate the *words* from *partwords* (three-syllable sequences that also occurred in the stream but spanned a word boundary). For each language, there were two test words and two a test partwords. For Language A, the test words were dakube and pogitae; the test partwords were baepagu and tedoki. For Language B, the test words were baepagu and tedoki; the test partwords were gitaeda and kubepo. As in Experiment 1, test words and partwords were synthesized and edited in the same way as described for the streams above, except that both were generated by having MacInTalk produce these three-syllable items in isolation. This produced a falling intonation on the final syllable of each item, making each test item (words and partwords) sound like a word spoken in isolation.

Within the 21-min stream of speech, each of the sixteen trisyllabic nonsense words occurred equally often, in a constrained order such that no word was immediately repeated and each word-final syllable could only be followed by either of two word-initial syllables. The transitional probabilities for non-adjacent segments inside words (between the vowels: segments 2, 4, and 6) were therefore 1.0; the transitional probabilities for non-adjacent segments within part-words were .5 (between segment 6 of one word and segment 2 of the next) and 1.0 (between segment 2 and segment 4). In order to discriminate between the test items, subjects would have to compute the transitional probability between non-adjacent segments, or another closely related statistic.

Test items did not differ in the frequency of individual segments or syllables. Importantly, they also did not differ in the transitional probabilities among adjacent segments or syllables. The languages were designed so that they could not be learned by computing transitional probabilities only among adjacent elements: as already noted, the adjacent transitional probabilities within and between words are all .5.

The apparatus and procedures were the same as in Experiments 1 and 2.

6.2. Results

Each animal's responses in each test condition (words vs. partwords) were converted to a percentage, and these percentages were averaged across animals. Fig. 5 presents the mean percent of trials showing an orienting response to words vs. partwords. Tamarins were significantly more likely to orient to partwords than to words (Wilcoxon Signed Ranks, $z(9) = -2.366$, $p = .018$ for Language A, $z(9) = -2.028$, $p = .0425$ for Language B, $z(18) = -3.107$, $p = .0019$ for A and B combined across each animal's useable test trials).

In short, then, while tamarins are not able to acquire patterns involving consonant segments (Experiment 2), they *are* able to acquire precisely comparable patterns involving vowels. This result also shows that the previous learning failures, for consonant patterns in Experiment 2, were not due to a general loss of interest by the tamarins in our synthetic speech materials or the complexity of those languages (which are, in almost every detail except the critical one, identical to the languages used in the present experiment). Apparently there is a true contrast in tamarins' ability to acquire the regularities of Experiments 1 and 3 and their inability to acquire the regularities of Experiment 2.

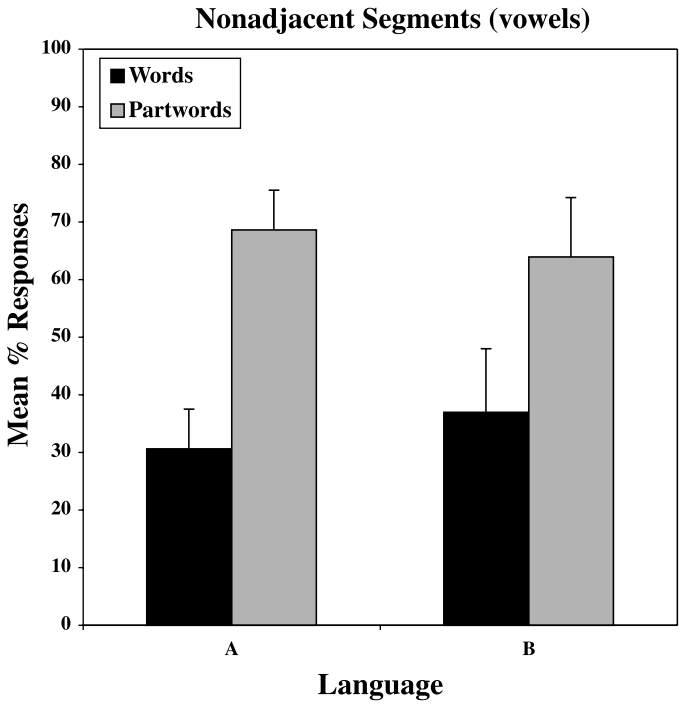

Fig. 5. Mean percent orienting responses to words versus partwords, for non-adjacent segment (vowel) languages (Language A on left, Language B on right). Data are from Experiment 3.

6.3. Discussion

The present results suggest that tamarins are able to acquire non-adjacent phonemic segment regularities, at least when the segments are vowels (which are loud and relatively long, as compared with stop consonants; see Liberman (1970), for a discussion of these contrasts in acoustic cues). Altogether, though, the pattern of results in tamarins is quite different from that of humans. In the section below, we summarize the results across these three experiments, compare them to our results with human adults, and suggest several importantly different hypotheses for what these cross-species results might mean.

7. General discussion

Our studies have investigated the abilities of tamarins to acquire certain fairly complex patterns in human speech stimuli, without any explicit training or reinforcement, but merely on the basis of listening and familiarization. The findings across our experiments show that tamarins, like humans, show very selective types of statistical learning: some of the languages we have constructed are consistently and easily learned,

while others are consistently not learned. However, a major finding of interest is that the selectivities of learning for tamarins are very different than those for human adults.

We began by asking whether tamarins (and, in Newport & Aslin, 2004, whether human adults) could acquire any statistical patterns built on non-adjacent regularities, or rather whether their statistical learning abilities would be limited to very elementary computations on immediately adjacent syllables. Our experiments clearly demonstrate that tamarins are not so limited: they are able to compute certain complex patterns involving non-adjacent elements. This important finding adds to the surprising inventory of abilities we have uncovered in this non-human species, including the ability to perceive and discriminate a set of human speech elements, keep track of the serial order of these elements and the probabilities that certain orders appear in a stream of speech, and in the present results also keep track of elements that regularly appear at some distance from one another. In recent (Fitch & Hauser, 2004) and future work we hope to extend these questions by asking whether tamarins can acquire other types of patterns, such as categories and hierarchies, that human languages display.

Surprisingly, however, the types of non-adjacent regularities that humans can readily acquire are not the same as those for tamarins. In Experiment 1, we found that tamarins are consistently able to acquire languages built from non-adjacent syllable regularities. However, human learners, tested on the same stimuli (as well as many other stimuli with the same type of structure), do not acquire this pattern, and human languages do not commonly exhibit this type of regularity (Newport & Aslin, 2004). In contrast, humans are readily able to acquire languages built from non-adjacent segment regularities, regardless of whether these regularities involve consonant patterns (with variation in the intervening vowels) or vowel patterns (with variation in the intervening consonants); and human languages often exhibit these types of patterns (see for example, Hebrew, Arabic, and Turkish; Newport & Aslin, 2004). However, tamarins do not appear to be able to acquire consonant patterns (see Experiment 2), even with repeated exposure and testing, though they can readily acquire vowel patterns (see Experiment 3). What do these differing results tell us about both tamarins and humans?

A first question is whether these results could arise from differences in the overall complexity of the languages or the differential availability of strategies based on computations among adjacent elements. As noted in Newport and Aslin (2004), the various non-adjacent languages we have constructed are carefully matched in many aspects of their overall complexity and are also designed to be similar to one another, and not readily acquired, in terms of statistical relations among adjacent elements only. All three types of languages utilize approximately the same inventory of sounds, are synthesized and edited in similar ways, and have comparable magnitude contrasts in the statistics that define words versus word boundaries. (In all cases the non-adjacent statistical regularities that define words are transitional probabilities of 1.0; those across word boundaries, and those between adjacent elements, are .33 to .5 in Experiment 1 and .5 in all other cases.) Moreover, the use of two language instantiations per language type (and the consistency of our findings across the two language instantiations, and often the repeated runs, within each experiment) insures that our results are due to the type of structure these languages exhibit, and not to the details of which

sounds are assigned to particular positions within the word. Finally, the different pattern of results across experiments was obtained with the same sample of subjects, using the same testing procedures.

One important question is whether there is any *adjacent* computational strategy by which tamarins could be learning our language materials. As already noted, all materials are structured so that adjacent transitional probability computations will not lead to learning the appropriate groupings. However, while we have described our languages primarily in terms of transitional probabilities, there are also differences between words and partwords in bigram or trigram co-occurrence frequency: that is, in the frequency with which these sets of two-syllable and three-syllable sequences occur in the corpus. While we have conducted a number of studies in human infants and adults to assess whether they compute transitional probabilities rather than co-occurrence frequency (Aslin et al., 1998; Fiser & Aslin, 2002; Hunt & Aslin, 2001), we have not directly studied this question in tamarins.[8] It is therefore possible that they are performing our tasks using trisyllabic frequency rather than transitional probabilities among elements. Interestingly, however, the patterns of learning in the present results suggest rather strongly that co-occurrence frequency is not the statistic computed by tamarins or humans in our tasks. If our learners were using trisyllabic frequency to analyze our speech streams, all of the languages we have utilized, in Experiments 1, 2, and 3, should be acquired in the same way. With respect to trisyllabic frequency, in all of the languages we have studied, words are more frequent than partwords. Therefore, if this were the computation performed by tamarins, they should be learning all of our languages equally well.[9] The selectivity of the results we have obtained argues that it is the type of statistic on which our languages differ—transitional probabilities among differing types of non-adjacent elements[10]—that tamarin and human learners are actually performing in the task.

Having argued, then, that there are no obvious complexity differences or alternative strategies that might readily explain our results, we now turn to several more

[8] In Aslin et al. (1998) we employed a design to control these co-occurrence frequencies and showed that human infants can still perform our task. This demonstrated that infants were capable of performing conditional probability computations rather than co-occurrence frequency computations. However, because of certain undesirable aspects of this design required to do such matching—individual syllable frequencies and trisyllabic frequencies of untested items are not matched—we do not use such a design in most of our studies. See Newport and Aslin (2004) for further discussion.

[9] An alternate possibility is that, if tamarins were utilizing trisyllabic frequency to learn our materials, they should acquire the non-adjacent syllable languages most easily (because there are only 6 words, that is, 6 trisyllabic sequences, in each of these languages), and the two types of non-adjacent segment languages exactly as well as one another (since each have 16 words, or 16 trisyllabic sequences). Indeed, in terms of trisyllabic sequences, the non-adjacent consonants languages and the non-adjacent vowels languages are virtually identical. The fact that tamarins acquire one type of non-adjacent segments language well (vowels), but do not acquire the other (consonants), argues strongly that they are not using this statistic for their learning.

[10] As noted earlier (see Footnote 1), while we have frequently used the term *transitional probability* to describe our materials, we mean more technically to refer to any of a class of conditionalized statistics, including *mutual information* or *conditional entropy*. The present point is that our results suggest the relevant statistic could not instead be a frequency statistic (such as tri-syllabic co-occurrence frequency).

theoretically interesting explanations: first, that there are differences in the types of elements that different species of learners readily perceive; and second, that there are differences among the learners in the way that non-adjacency in speech is handled.

7.1. Species differences in basic computational elements: Syllables, consonants, and vowels

One possible difference between human adults and tamarin monkeys is in the types of units they perceive in human speech and/or the types of units on which they are able to perform statistical computations. In order to acquire non-adjacent syllable regularities, listeners must be able to perceive syllables as units in the speech stream and then keep track of which syllables occur in which order, how frequently, and the like. In contrast, in order to acquire non-adjacent segment regularities, listeners must perceive and compute statistics on consonants and vowels. A possible explanation of our tamarin findings, then, is that tamarins are capable of perceiving (and computing statistics on) syllables and vowels—both relatively long, loud, and acoustically prominent units in a rapid speech stream—but not consonants. A variety of difficulties with consonants could lead to failures in acquiring consonant patterns: they might not perceive consonants (particularly stop consonants) as separable units from their subsequent vowels; they could be unable to categorize the different instances of a particular consonant as the same phoneme when the consonant appears with different vowels; or they could be unable to compute and maintain the statistical properties of consonants over a lengthy exposure. Any of these difficulties would be consistent with the complex acoustic/phonetic properties of stop consonants, as compared with vowels or full syllables, that have been observed in the human speech perception literature (see Liberman, 1970, for discussion).

None of these abilities has previously been tested in tamarin monkeys, so any of them might account for our results. However, there is important evidence on some of these abilities in other non-human species. Kuhl and Miller (1975) demonstrated that chinchillas could be trained to respond to discriminate between sets of CV syllables that differed in their initial consonants (/d/ versus /t/), even in the face of varying vowels (e.g., ti, ta, tu versus di, da, and du), and moreover could generalize this discrimination to syllables with the same consonants but new vowels produced by the same talkers. Some of these abilities have also been shown in macaques (cf. Kuhl, 1989, for more general discussion of these issues). Nonetheless, even these results do not indicate whether non-human listeners can extract and categorize consonants from running fluent speech, rather than in isolated CV syllables, or do so without reinforcement training. Most relevant to the present task, we do not know whether non-humans can compute and maintain over a lengthy exposure the statistical properties of consonants, separately from the statistics of surrounding vowels or the syllables in which they are embedded. The greater acoustic prominence and perceptual salience of vowels and syllables might make them easier to process and to utilize in computations.

If this explanation is correct, would we suggest that *human* listeners can only perceive or compute the statistical regularities of segments (both consonants and vowels), but not syllables? This is in fact a possibility, given the human results we have

obtained. We would not argue, however, that human listeners perceive only segments and not syllables; extensive evidence in the psycholinguistics literature suggests that human listeners can perceive and respond to both syllables and segments, depending on the task (cf. Nygaard & Pisoni, 1995, for a review). For example, humans can perform both phoneme and syllable monitoring, and can learn to read both syllabary and alphabetic scripts. But it is possible that human listeners perform statistical computations at only one of these levels of representation (segments), or that they perform their computations initially on segments and construct syllable statistics indirectly, through segment combinations. This is one of the accounts we hypothesized for the human data (Newport & Aslin, 2004). As noted in that paper, our previous studies of statistical learning (Saffran, Aslin & Newport, 1996; Saffran, Newport & Aslin, 1996) involved words that were formed from high transitional probabilities among both neighboring syllables *and* neighboring segments. We described our materials in terms of syllable statistics, for simplicity of presentation, but the design was not intended to address the question of whether our subjects were, in fact, performing their computations on the syllables or on the segments of our speech streams. In contrast, our new materials involving non-adjacent syllable and segment regularities each require one of these types of computation and are controlled on the other. A possible interpretation of the present results from humans is that adult human learners perform their computations exclusively on segments and are unable to acquire statistical groupings based on syllables.

An important question for future study is how human infants perform on these types of problems, and therefore what type of units they might use to perceive speech and compute its regularities in the early stages of language acquisition. One possibility is that human infants are, from the beginning, like human adults, perceiving speech in terms of phonetic segments (consonants and vowels) and keeping track of the regularities of speech initially in terms of these units (Hillenbrand, 1983, 1984; cf. Kuhl, 1985, 1989, for discussion). Another possibility, suggested on the basis of findings with very young infants (Bertoncini & Mehler, 1981; Jusczyk & Derrah, 1987; cf. Jusczyk, 1997, and Mehler, 1985, for discussion), is that infants begin perceiving speech in terms of larger and more holistic units, such as syllables, and only later—either through maturation or through learning—analyze those syllables into the features and segments thought to characterize speech representations in adults. Ongoing research using our non-adjacent syllable and segment materials with 8- and 10-month-old infants suggests that they cannot perform any of these non-adjacent computations, and may be limited to computing only adjacent regularities. In order to address the question of perceptual and computational units in adults and infants, then, we are in the process of designing a new set of studies that contrast syllables versus segments without involving non-adjacency.

7.2. Species differences in non-adjacency

A second alternative for explaining our species differences concerns non-adjacency itself, and how non-adjacent regularities are computed by tamarins versus human

adults. Our findings suggest that non-adjacency per se is not a problem for tamarins: they are able to compute both non-adjacent syllable regularities and non-adjacent segment (vowel) regularities; only consonants give them difficulty, likely arising from the acoustic complexities of consonants themselves. What is surprising, then, is that certain types of non-adjacency are extremely difficult for human listeners.

Why should non-adjacency—particularly syllable non-adjacency—be difficult for human listeners and relatively easy for tamarin monkeys? As noted above, this is not likely to be because tamarins are in general more cognitively capable than adult humans. It must therefore be because human speech is processed in a different way by humans than by tamarins, and particularly in such a way that the computation of non-adjacent syllable regularities becomes more complex for human adults. We suggest three reasons why this might be true.

First, as discussed in the previous section, human adults might process or compute regularities primarily in terms of segments, especially in these kinds of tasks; syllable regularities might then be inaccessible, or might be available only indirectly, by assembling them from segment regularities. If this interpretation is correct, then keeping track of *non-adjacent syllable* regularities would be extremely indirect and fairly complicated. In contrast, if tamarins hear human speech in terms of larger, more holistic elements like syllables, keeping track of syllable regularities—whether adjacent or non-adjacent—would be relatively easier.

Second, human adults and tamarin monkeys might differ in the computational explosion problem we described in our introduction to non-adjacent regularities (see Newport & Aslin, 2004). A general problem of inductive learning is the large number of hypotheses that might describe any set of patterned materials. In statistical learning, this problem translates into an explosion in the number of computations a learner might in principle perform. Given a lengthy speech stream, a learner that was restricted to perceiving and remembering only adjacent syllables would have a fairly big job on his hands; but a learner that could perceive and compute all the adjacent and non-adjacent segment and syllable relations over a 21-min stream of speech (or even a 1-min stream of speech) would have an intractable problem. For this reason we have suggested (cf. Newport & Aslin, 2004) that human adults might sensibly have great difficulty acquiring non-adjacent syllable regularities, where there is nothing in the non-adjacent syllables to bring their patterned relationships to the attention of the learner. In order to find non-adjacent syllable regularities in this type of stream, a good learning device would have to keep track of the regularities among all adjacent syllables, all syllables one away, all syllables two away, etc., before finding where the patterned regularities lie.[11] In contrast, non-adjacent segment regularities among segments of the same type—either consonants or vowels—and skipping over segments of the opposite type would be relatively easier to perceive and pose a relatively easier and more restricted computational problem. Indeed, if consonants and

[11] On this view, the success of infants and adults in the paradigms of Marcus, Vijayan, Bandi Rao, and Vishton (1999) and Gomez (2002) is perhaps made possible by the fact that the strings are already segmented and are only three syllables long. In such restricted materials, the number of potential adjacent and non-adjacent regularities is very small.

vowels are represented on separate 'tiers,' as hypothesized by many phonological theories, these regularities do not involve non-adjacency at all (see Newport & Aslin, 2004, for further discussion).

On this account, how could tamarins solve a computational problem that exceeds the capacities of adult humans? It is possible that, in this problem as in certain other learning problems (Elman, 1993; Goldowsky & Newport, 1993; Newport, 1988, 1990), 'less is more.' Suppose that tamarins have a greatly restricted ability, compared with human adults, to attend to a lengthy and complicated stream of human speech. While we play our materials for 21-min, continuously, to tamarins, we do not know how they are attending to this stream and where the crucial computations are being performed. One possibility is that tamarins might be attending for a few syllables or a slightly longer stretch, then tuning out, later tuning in again, and so forth. (Human adults might similarly be tuning in and out, particularly since our speech streams are long and extremely boring. But on the present hypothesis, if humans attend for longer stretches than tamarins, however long these stretches are, their computational problem is more severe.) Such a restriction on the length of stretches in which adjacent and non-adjacent computations are performed will result in fewer required computations, but will still capture those that are highly patterned, as long as the patterns appear within relatively short ranges (see Goldowsky & Newport, 1993, for a more detailed explanation of this argument, though applied to a different learning problem).

In other words, for regularities among elements that are at moderate distances from one another, a cognitively limited learner might be more capable than a learner with greater cognitive capacities. This hypothesis also predicts that an appropriately aged infant might be like a tamarin, though this prediction depends on many other factors, such as whether infants at various ages hear speech in terms of syllables at all, and whether they have sufficient processing capacities to acquire any non-adjacent patterns. Ongoing developmental work examines this question further.

A third and final possibility is that the concepts of 'distance' and 'non-adjacency' differ greatly, depending on the representations a learner has of the materials to be learned. This possibility encompasses some of the issues already raised above, but with a somewhat different emphasis. If tamarins hear human speech as relatively large, holistic chunks of sound, occurring in a linear sequence, then the acquisition of non-adjacent syllables reduces to keeping track of those syllables and how often they occur together; this problem is only slightly more difficult than acquiring adjacent syllables, in that all the syllables one away would have to be computed. In contrast, the much more complex and well-articulated representation that adult humans have of human speech might make non-adjacent syllables much more distant from one another than they are for tamarins.

Human adults are thought to perceive and represent speech syllables as comprised of features and (perhaps) segments that combine in a hierarchical fashion. For humans, speech sounds are not holistic chunks that are ordered in a linear sequence (Liberman, 1970). Rather, sound elements are organized hierarchically, with various types of elements combining to form higher order elements. The syllable is thought to consist of an onset consonant plus a rhyme; the rhyme consists of a nucleus vowel plus a possible final consonant. Syllables are then grouped together into metrical feet (on

the phonological side) and stems and affixes (on the lexical side). It is not entirely clear yet how to think about these types of highly differentiated and articulated representations in an understanding of statistical learning. Do adults (and infants) conduct their statistical computations on a finely articulated and hierarchically organized set of representational units? If so, non-adjacent syllables might be hierarchically quite distant from one another, much farther apart than in a simpler tamarin-like representation. An understanding of the details of such a process will require much further experimentation, to determine more precisely what types of units are used for computing and how linear versus hierarchical distance affects statistical learning.

8. Conclusions

We have suggested two important types of conclusions from our findings in these experiments. First, we have shown that tamarin monkeys have a number of capabilities for perceiving human speech and computing its regular patterns, including very surprising ones not previously demonstrated. In order to perform our learning tasks, tamarins must perceive human speech syllables and vowels, and must be able to keep track not only of the regular sequences in which they occur, but also the regular non-adjacent patterns. Tamarins do not appear to be readily capable of learning patterns among successive consonants separately from vowels; and we expect that there will be other more complex patterns they will be unable to learn. At present, however, our studies continue to reveal unexpected and surprising abilities to process fairly complex aspects of human speech in a non-human species, in the absence of any explicit training.

Finally, we believe the contrasts—both the similarities and the differences—between tamarins and human adults on our tasks provide important insights into the nature of speech perception and statistical learning. As previous literature has argued, studies of non-human listeners provide unique evidence regarding the generality of human speech abilities and the sources from which they have become specialized (Hauser, 1996; Hauser et al., 2002; Kuhl, 1986, 1989; Trout, 2000). Perhaps most significantly, we believe the contrasts between tamarins and humans in our studies suggest that statistical learning may be an extremely interesting, complex and well-articulated learning mechanism—one that is shared across species and domains to some extent, but is in certain ways constrained and differentiated across species and domains. The present evidence suggests that, while tamarin monkeys and human adults both perform statistical computations on streams of speech and can use these computations to segment the streams into coherent groups of sounds, they may not be performing their computations on the same types of elements or assembling the outcomes in the same ways. The simplest difference we have suggested is in terms of the elements on which computations are performed: Tamarins may be performing their statistical computations only on large, holistic speech elements, such as syllables and vowels. Humans, in contrast, appear to be performing their statistical computations on more finely differentiated speech elements, such as consonants and vowels, and perhaps not on syllables. A second possibility is that

tamarins may differ from humans in the length of portions of the stream over which they perform statistical computations. Given the computational explosion problem, such a difference may in turn create differences in the types of regularities that are easy and difficult to acquire. A final possibility is that tamarins and humans may differ in the way they represent relations among speech elements, and therefore in what kinds of elements become proximal or distant from one another. Tamarins may represent speech as a linear sequence of sounds elements, such as ABC. On this type of representation, elements A and C are just one element apart. In contrast, humans are thought to represent speech as comprised of hierarchical structures among a number of levels of inter-related elements. If statistical learning is performed on such a structured representation—and not merely on surface strings of elements—then A and C in the above example have a more indirect relation to one another and might be much farther apart computationally.

Further research will be required to determine which of these interpretations is correct. In the meantime, our results suggest that statistical learning is not just a simple, rapid, but very limited or elementary process. The ability to learn using statistical computations appears over a wide range of patterns and types of materials, and across a number of types of learners. But the present evidence suggests strongly that the way these computations are done depends on the types of perceptual representations the learner brings to the task. We hope in future research to reveal in more detail how tamarin monkeys, human adults, and human infants employ this mechanism in learning, and how, possibly, the unique nature of human languages may have arisen from unique forms of learning in humans.

Acknowledgments

We are grateful to Kelly Kinde and Joanne Esse for their help in constructing the stimuli for these experiments, and Robb Ruttledge for help in testing the tamarins. This research was supported in part by NSF Grant SBR-9873477 to RNA, ELN, and MDH, NIH Grant DC00167 to ELN, and NIH Grant HD37082 to RNA.

Appendix A. Human performance on tamarin stimuli and comparison with tamarin performance on the same stimuli

As noted in the Method Sections of each experiment, the stimulus and testing materials used with tamarins in this paper and those with human adults in Newport and Aslin (2004) were identical in all ways except speech rate (the stimuli for tamarins were slowed down by 10%) and, in some cases, the particular partwords used as test items (the smaller number of test items used with tamarins required choosing different partwords in order to cover the range of possible partwords equally well). In order to be certain that our obtained differences and similarities between tamarins and humans were indeed species differences (and not due to these small differences in materials), we ran additional human subjects using the exact stimulus and testing

materials used with tamarins. The results were precisely like those reported in Newport and Aslin (2004) with the original materials, and are presented below. As reported in Newport and Aslin (2004), human adults did not show learning of the Non-adjacent Syllable languages (materials from Experiment 1 of the present paper), but they did show highly significant learning of the Non-adjacent Segment languages involving both consonants (materials from Experiment 2 of the present paper) and vowels (materials from Experiment 3 of the present paper). This is in contrast with tamarins, who showed learning in Experiments 1 and 3, but not in Experiment 2. In order to compare these results across species directly, we also list below the d' measures for humans and tamarins on the same materials. These results demonstrate that the contrast in our findings across species is not due to the small differences in stimulus materials across the two series of experiments.

It does, of course, remain possible that the species differences arise from differences in the methodologies used to obtain responses from human adults vs. tamarins. (Human adults were asked to choose a word versus a partword on a two-alternative forced choice test. Tamarins were presented with a word or a partword on a series of test trials and were scored for spontaneous looking toward the speaker.) It seems extremely unlikely, however, that these methodological differences could explain the systematic differences in patterns of learning that we obtained across species.

Human performance on tamarin stimulus materials				Tamarin performance
Nonadjacent syllables (stimuli from Exp. 1)				
Choice of words vs. partwords on 2AFC			d'	d'
	% Correct			
A	38.54	$t(11) = -1.77$, NS	−.32	1.08
B	43.75	$t(11) = -.80$, NS		
Nonadjacent segments—consonants (stimuli from Exp. 2)				
Choice of words vs. partwords on 2AFC			d'	d'
	% Correct			
A	80.00	$t(4) = 4.95$, $p < .01$	1.25	.03
B	82.50	$t(4) = 3.56$, $p < .05$		
Nonadjacent segments—vowels (stimuli from Exp. 3)				
Choice of words vs. partwords on 2AFC			d'	d'
	% Correct			
A	96.25	$t(4) = 18.50$, $p < .0001$	1.92	.84
B	86.25	$t(4) = 2.96$, $p < .05$		

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.

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant perception. *Infant Behavior and Development, 4*, 247–260.

Chomsky, N. A. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chomsky, N. A. (1995). *The minimalist program*. Cambridge: MIT Press.

Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world: Inside the mind of another species*. Chicago: Chicago University Press.

Cleveland, J., & Snowdon, C. T. (1981). The complex vocal repertoire of the adult cotton-top tamarin, *Saguinus oedipus oedipus*. *Zeitschrift fur Tierpsychologie, 58*, 231–270.

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

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.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*, 377–380.

Ghazanfar, A. A., Smith-Rohrberg, D., Pollen, A., & Hauser, M. D. (2002). Temporal cues in the antiphonal long calling behaviour of cotton-top tamarins. *Animal Behaviour, 64*, 427–438.

Goldowsky, B. N., & Newport, E. L. (1993). Modeling the effects of processing limitations on the acquisition of morphology: The less is more hypothesis. In J. Mead (Ed.), *The proceedings of the 11th West Coast Conference on Formal Linguistics*. Stanford, CA: CSLI.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.

Greenfield, P. M. (1991). Language, tools, and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences, 14*, 531–595.

Hailman, J. P., & Ficken, M. S. (1987). Combinatorial animal communication with computable syntax: Chick-a-dee calling qualifies as 'language' by structural linguistics. *Animal Behaviour, 34*, 1899–1901.

Hauser, M. D. (1996). *The evolution of communication*. Cambridge: MIT Press.

Hauser, M. D. (1997). Artifactual kinds and functional design features: What a primate understands without language. *Cognition, 64*, 285–308.

Hauser, M. D. (1998). Expectations about object motion and destination: Experiments with a nonhuman primate. *Developmental Science, 1*, 31–38.

Hauser, M. D. (2000). *Wild minds: What animals really think*. New York: Henry Holt, Inc.

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. D., Dehaene, S., Dehaene-Lambertz, G., & Patalano, A. L. (2002). Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (*Saguinus oedipus*). *Cognition, 86*, B23–B32.

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

Hauser, M. D., Weiss, D., & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition, 86*, B15–B22.

Hillenbrand, J. (1983). Perceptual organization of speech sounds by infants. *Journal of Speech and Hearing Research, 26*, 268–282.

Hillenbrand, J. (1984). Speech perception by infants: Categorization based on nasal consonant place of articulation. *Journal of the Acoustical Society of America, 75*, 1613–1622.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Simultaneous extraction of multiple statistics. *Journal of Experimental Psychology: General, 130*, 658–680.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge: MIT Press.

Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology, 23*, 648–654.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*, 69–72.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America, 63*, 905–917.

Kuhl, P. K. (1985). Categorization of speech by infants. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming buzzing confusion*. Hillsdale, NJ: Erlbaum.

Kuhl, P. K. (1986). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology, 45*, 233–265.

Kuhl, P. K. (1989). On babies, birds, modules, and mechanisms: A comparative approach to the acquisition of vocal communication. In R. J. Dooling & S. H. Hulse (Eds.), *The comparative psychology of audition: Perceiving complex sounds*. Hillsdale, NJ: Erlbaum.

Kuhl, P. K. (1991). Human adults and human infants show "a perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93–107.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology, 1*, 301–323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431–461.

Lieberman, P. (1984). *The biology and evolution of language*. Cambridge: Harvard University Press.

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

Mehler, J. (1985). Language related dispositions in early infancy. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming buzzing confusion*. Hillsdale, NJ: Erlbaum.

Miller, C. T., Miller, J., Gil-da-Costa, R., & Hauser, M. D. (2001). Selective phonotaxis by cotton-top tamarins (*Saguinus oedipus*). *Behaviour, 138*, 811–826.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63*, 176–185.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition. *Language Sciences, 10*, 147–172.

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

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

Nygaard, L. C., & Pisoni, D. B. (1995). Speech perception: New directions in research and theory. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication*, in: E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception and cognition*. 2nd ed. San Diego: Academic Press.

Ramus, F., Hauser, M. D., Miller, C. T., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and cotton-top tamarin monkeys. *Science, 288*, 349–351.

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

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

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.

Santos, L. R., & Hauser, M. D. (1999). How monkeys see the eyes: Cotton-top tamarins' reaction to changes in visual attention and action. *Animal Cognition, 2*, 131–139.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development, 58*, 1–221.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart & Winston.

Stebbins, W. C. (1983). *The acoustic sense of animals*. Cambridge: Harvard University Press.

Trout, J. D. (2000). The biological basis of speech: What to infer from talking to the animals. *Psychological Review, 108*, 523–549.

Weiss, D. J., Garibaldi, B. T., & Hauser, M. D. (2001). The production and perception of long calls in cotton-top tamarins (*Saguinus oedipus*): Acoustic analyses and playback experiments. *Journal of Comparative Psychology, 15*, 258–271.

Weiss, D. J., & Hauser, M. D. (2002). Perception of harmonics in the combination long call of cotton-top tamarins (*Saguinus oedipus*). *Animal Behaviour, 64*, 415–426.

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour, 63*, 293–299.