Cognitive Science 33 (2009) 1087–1116

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01041.x

Changing Structures in Midstream: Learning Along the Statistical Garden Path

Andrea L. Gebhart, Richard N. Aslin, Elissa L. Newport

Department of Brain and Cognitive Sciences, University of Rochester

Received 3 October 2008; received in revised form 19 February 2009; accepted 22 February 2009

Abstract

Previous studies of auditory statistical learning have typically presented learners with sequential structural information that is uniformly distributed across the entire exposure corpus. Here we present learners with nonuniform distributions of structural information by altering the organization of trisyllabic nonsense words at midstream. When this structural change was unmarked by low-level acoustic cues, or even when cued by a pitch change, only the first of the two structures was learned. However, both structures were learned when there was an explicit cue to the midstream change or when exposure to the second structure was tripled in duration. These results demonstrate that successful extraction of the structure in an auditory statistical learning task reduces the ability to learn subsequent structures, unless the presence of two structures is marked explicitly or the exposure to the second is quite lengthy. The mechanisms by which learners detect and use changes in distributional information to maintain sensitivity to multiple structures are discussed from both behavioral and computational perspectives.

Keywords: Statistical learning; Word segmentation; Primacy; Representation

1. Introduction

One of the challenges faced by any robust learning mechanism is to decide whether a given sample of exemplars is best characterized by a single pattern or by multiple patterns, characteristic of different subgroups of exemplars. A second important challenge faced by a learning mechanism is to determine when these patterns have undergone a change—that is, when the exemplars are no longer arising from the same underlying source. Although human learners may have a stationarity bias—a prior bias that structures do not undergo rapid and frequent changes—the inability to detect structural changes and maintain more

Correspondence should be sent to Andrea L. Gebhart, Department of Brain and Cognitive Sciences, Meliora Hall, River Campus, University of Rochester, Rochester, NY 14627-0268. E-mail: agebhart@bcs.rochester.edu

than a single representation would be a serious impediment to efficiently learning the complex and dynamic structures that characterize any learning problem. The goal of the present series of experiments is to address this second challenge in the domain of auditory statistical learning: How do human learners extract the patterns from a corpus of input that undergoes a structural change, and what cues in the input successfully trigger such a representational change?

1.1. Cues for detecting changes in structure

Two potential sources of information could lead a learner to detect that there has been a change in structures over time. The most obvious is a contextual cue, such as the day of the week or a sound in the distance, that correlates with (or signals) a change in the structure of the input set (Alloy & Tabachnik, 1984). An ideal learner could then perform computations over a subset of exemplars from each context and determine whether the two structures differed by some criterion. The problem with this strategy is that there is a very large (in principle, infinite) number of possible contexts and contextual cues. However, if human learners ignored all contextual variables, they would make many errors, such as assuming that schools are open every day with probability .536 (5/7 days per week × 9/12 months per year). Thus, learners must make implicit decisions about when and what types of contextual cues are relevant in signaling a change in structures.

A second potential source of information about a change in structure comes from monitoring the consistency of the structure itself (Basseville & Nikiforov, 1993; Gustafsson, 2000). If a random sample of size N always provided the learning mechanism with the same distributional information, then the learner would conclude that the structural information is uniform. Any deviation from the past structural representation (by some criterion) upon encountering the next sample of size N would indicate that the structure had undergone a change. Selection of N involves a trade-off: If N is too small, then detection of a true change in structure will be unreliable; if it is too large, then any change will be detected more slowly than is optimal.

An ideal learner could compute the variance over subsets drawn from a very large corpus of exemplars and determine the minimal subset size that yields asymptotically low variance (Kareev & Fiedler, 2006). But human learners are notoriously impatient and draw tentative conclusions about structures from rather small samples (Kareev, Lieberman, & Lev, 1997; Tversky & Kahneman, 1971). However, if learners base their structural hypotheses on small samples, it opens the door to garden-path errors, and most learning models have difficulty recovering from such errors without huge amounts of countervailing data to overcome the initial incorrect structural hypothesis (Hogarth & Einhorn, 1992).

1.2. Computational approaches to detecting changes in structure

Detecting a change in the structure of a sequential dataset is a common problem in many domains of science and engineering. Consider the example of astronomical time-series data in which the goal is to find evidence that a catastrophic event has occurred. Such an event

must be discriminated from random fluctuations in the background activity of the dependent measure (e.g., radio waves). Scargle, Norris, and Jackson (2006) describe a computational approach called Bayesian Blocks that partitions the time series into bins (either of fixed duration or a sliding window) and searches for discontinuities at a bin boundary. The structure before the discontinuity is relatively stable, and then it undergoes a transition into a different structure after the discontinuity.

In the foregoing example, the entire time series is accessible in batch mode to find the optimal fit to the structural change. Moreover, only a single change is expected against a stable background. A more general solution to the problem of change detection is a *mixture of experts* approach (Jacobs, Jordan, Nowlan, & Hinton, 1991) in which two or more neural networks adaptively partition the input corpus based on a competitive-learning process, and each network learns a different subset of this corpus. In this case, the corpus may be partitioned into many subsets. A drawback of this approach is that the entire corpus is partitioned iteratively (i.e., in batch mode) to find a good fit. However, the general mixture models approach can also be implemented in a single pass to fit a time series as it unfolds incrementally (Kehagias & Petridis, 1997). What remains unclear is how well such an unsupervised mixture model fits data on human learning. That is, is the performance of human learners who are confronted with an input corpus containing sudden (perhaps multiple) changes in structure well described by a mixture model?

1.3. Empirical evidence of detecting changes in structure

Studies of learning have provided extensive evidence for the limits of detecting two structures. In studies of classical and operant conditioning in nonhuman animals, there is evidence for both forward and backward blocking, in which the effectiveness of a compound conditioned stimulus is reduced by prior or subsequent presentation of a single component of that compound (Miller, Barnet, & Grahame, 1995). These blocking effects are also shown by human adults (Shanks, 1985) and by preschool-aged children (Sobel, Tenenbaum, & Gopnik, 2004). It is important to note that, in order for such blocking effects to occur, learners must be treating the sequential parts of their learning experiences as arising from a single continuous environment, with learning from the earlier portion of exposure strongly affecting that from the second; they are not separating them, as an ideal learner would, when the environment was determined to change. In adult verbal learning studies, a switch from one set of paired associates to a second partially overlapping set leads to the rapid extinction of the initially learned pairings (Barnes & Underwood, 1959). Studies of contingency learning in human adults find both primacy effects and recency effects (Marsh & Ahn, 2006), with primacy effects supportive of Tversky and Kahneman's (1974) notion of anchoring. Finally, in the domain of language learning there is ample evidence for primacy effects that interfere with later learning of new structures, both at the level of phonetics and phonology (see Werker & Curtin, 2005) and at the level of morphology and syntax (see Johnson & Newport, 1989). Thus, exposure to two successive structures yields a diverse set of learning outcomes, depending on the species, age, domain, and task.

1.4. Potential learning outcomes with exposure to sequential structures

The foregoing computational and empirical reviews highlight the uncertainty about which of four possible outcomes of learning two different successive structures (A and B) will be obtained in a given domain: learn A but not B, learn B but not A, learn both A and B, or learn neither A nor B. Two additional literatures provide hints that learning an initial structure can interfere with the learning of a second structure.[1] Junge, Scholl, and Chun (2007) used a visual search task in which the spatial location of a target was predicted by the context of the nontarget distractor set. Prior studies had shown that this form of contextual-cueing facilitated reaction times to search for the target. However, Junge et al. showed that improvements in reaction time due to contextual cueing were eliminated if spatial cueing was preceded by random cueing. This suggests two possibilities. The first is that the participants are learning that the first input set is random and continue to expect randomness, which competes with the later structure. The second is that they have learned ''false structures'' that are attested in the initial random input set (yet are rare because of sparse data), and that these false structures compete with the ''real structures'' that follow in the second input set.

Studies of artificial grammar learning (AGL) also provide evidence relevant to the learning of two structures. In the AGL literature the question typically investigated is whether learners show generalization across two different inventories of elements when the structure remains the same (e.g., Lany, Gomez, & Gerken, 2007). However, Conway and Christiansen (2006) (Experiment 3b) asked whether learners could extract two different structures, each of which had its own inventory of elements. In contrast to the present series of experiments, Conway and Christiansen alternated the structures repeatedly rather than presenting each structure only once. They reported that one of the two structures was learned while the other was not, despite the obvious cue provided by the two different inventories of elements. However, given the design, it was not clear if the first structure was the only one that was learned, despite subsequent structural alternations.

Finally, Weiss, Gerfen, and Mitchel (2009) used two streams of speech composed of trisyllabic nonsense words that alternated repeatedly every 2 min. The gender of the talker differed between the two alternating speech streams, thereby providing a strong contextual cue for the change in structure. When the structures of the two streams were different, adults were able to learn them both; in the absence of the contextual (talker) cue, learning was at chance for both.

1.5. The present series of experiments

The goal of the present series of experiments was to explore the mechanisms that enable learners to detect a change in underlying structure in the speech domain using the well-studied paradigm of statistical learning—a robust and rapid form of distributional learning that has been studied in the auditory, visual, and tactile modalities (see reviews by Aslin & Newport, 2008; Perruchet & Pacton, 2006; Saffran, 2003). In contrast to Weiss et al. (2009), we are not focused on bilingualism, which is an interesting example involving the learning of two structures with strong contextual cues (e.g., differences in talker and phonetic

inventory). Rather, we consider the more general case of a corpus of input that could arise from a single structure or from more than one structure, changing through time. In the absence of strong contextual cues, how does the learner determine the number of structures given the *variance* in the structural estimates that naturally arise from small samples?

Our focus here is on auditory statistical learning using artificial speech streams composed of consonant-vowel (CV) syllables. Our experimental paradigm for auditory statistical learning studies (see Aslin & Newport, 2008) consists of an exposure phase, during which adult learners extract sequential patterns from a continuous stream of synthesized speech, followed immediately by a test phase, in which learners judge the relative familiarity of fully consistent patterns (''words'') versus partially consistent patterns (''part-words'') that were embedded in the exposure stream. The patterns used to create the original speech stream are three-syllable nonsense words concatenated using a speech synthesizer so that only statistical regularities and no other cues (such as pauses, immediate repetitions, or prosody) are available to the learner to extract the inventory of words.

For practical reasons, we chose a particular experimental design to explore how learners become sensitive to a change in structure. The statistical structures contained in the input of our previous studies were uniform across the entire exposure corpus. That is, each block of words within the corpus of a language had precisely the same distributional properties, albeit in different randomized orders.

In the present experiments, we first assessed performance on each of two individual structures in separate control groups. We verified that learners achieved excellent performance on each individual structure. Then, for separate sets of participants, we introduced a step-change in structure after the first structure. In our main experiments (Experiments 1b and 2b), this shift was not signaled, in order to investigate whether the mere change in pattern would provide an adequate cue to learners that would shift their performance. We then tested learners after the second structure had been presented for a duration that was sufficient for learning. A key design feature of our experiments was the elimination of any aggregate statistics across the two learning phases that could enable a learner who failed to separate the data to achieve above-chance performance on the postlearning test phase. In a final posttest, we asked whether learners had acquired the structural information present in both the first and second learning phases. If they did, then they must have been monitoring the input streams effectively. If they did not learn the second set of structures, then either they failed to monitor the input within the durations of presentation of the two learning phases, or learning during the first phase interfered with learning during the second phase. We answered this question by comparing performance in these studies to that in the control studies (in which participants had been exposed to only one structure).

To address these questions we produced two languages that could each be learned from fairly short exposure phases. Each of the two languages consisted of a set of 16 nonsense words that were easily segmented from fluent streams of speech in a previous study (Newport & Aslin, 2004), and we verified in the present study that robust learning of each of the languages ($\sim$80% correct) could be accomplished after 5 min of exposure. Although this did not allow us to assess the time course of learning continuously, it did limit the amount of exposure for each language to a total of 480 word tokens (30 tokens of each

word). We then presented the exposure phases of the two languages in sequence, in a 10-min stream of speech that was either marked, or not marked, by a cue signaling the change in structure. At the end of the second exposure stream, participants were presented with two-alternative forced choice (2AFC) test trials that compared words and part-words in the first language and also words and part-words in the second language. These intermixed test trials assessed whether participants learned the words of the first language, the second language, both languages, or neither language.

Most statistical learning tasks have presented structures that are noise-free; that is, all of the elements in the inventory join with one or more other elements to form higher-order structures (see Hudson Kam & Newport, 2005, and Shukla, Nespor, & Mehler, 2007, for exceptions). Such signal-only learning tasks are, of course, artificial and unlikely to be representative of natural environments that contain both structured (signal) and unstructured (noise) elements. This raises the following dilemma for the learner, as noted by Bialek, Nemenman, and Tishby (2001): ''the structure or order in a time series or a sequence is related almost by definition to the fact that there is predictability along the sequence.... [but] the predictive information is a small fraction of the total information. Can we separate these predictive bits from the vast amount of nonpredictive data?'' (p. 2453). This task of learning what is signal and what is noise is an intrinsic part of determining when the input has undergone a change in structure.

In the present series of experiments, we presented adult participants with two successive, structured input sets (''languages'') and varied a number of contextual variables to assess the interactive effects on learning. We chose not to use random input sets because current measures obtained with implicit learning paradigms do not allow us to determine whether participants have learned that the input set is random or whether they have learned false structures. By using two different structures we can assess what has, or has not, been learned from each. Moreover, by using speech streams we can determine whether the primacy effects observed by Junge et al. (2007) are specific to a contextual-cueing task in the visual domain. Importantly, in contrast to Conway and Christiansen (2006) and Weiss et al. (2009), we introduced only a single transition between the two structures rather than repeated alternations. Our goal was to focus on the detection of the change in structure and the effects of that change on maintenance or loss of sensitivity to the first structure.

2. Experiment 1a: Language A or Language B alone

In our previous studies (Aslin, Saffran, & Newport, 1998; Newport & Aslin, 2004; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996), we have shown that human learners can acquire many different types of patterned relations among syllables or segments by rapidly computing a set of statistics that indicate how consistently the syllables or segments occur together in the speech stream.

The goal of the present experiment was to select and pretest two languages, out of the many languages used in our earlier studies, each of which was: (a) learned well on its own with 5 min of exposure, and (b) learned to approximately the same performance level as the

other language by separate groups of participants. Languages that met these criteria could then be used in subsequent experiments investigating our main questions. We selected two languages with a vowel-frame structure that had been used by Newport and Aslin (2004) and learned quite well in a 20-min exposure, and then we verified in the present experiment that both of these languages met the foregoing requirements.

2.1. Method

2.1.1. Participants

Sixteen undergraduates, who were enrolled at either the University of Rochester or Monroe Community College, participated in the present experiment for a payment of $10 each. In this and the following experiments, all participants were monolingual native English speakers, and none had previously participated in a statistical learning experiment. In addition, in this and the following experiments, all participants reported that they had normal hearing and that they had no known learning disabilities or attention disorders. Eight participants were assigned to Condition 1 (5 min Language A exposure), and eight participants were assigned to Condition 2 (5 min Language B exposure).

2.1.2. Design and materials

Table 1 illustrates the design and composition of each language. Each language used the same inventory of six consonants (b, p, d, t, g, k) and six vowels (a, i, u, e, o, ae). Both languages were structured so that the vowels formed a consistent word-frame while the consonants varied. Each language consisted of 16 trisyllabic words, constructed from two unique three-vowel frames with two different consonants possible in each of the consonantal positions. The transitional probabilities between the vowels within a word were 1.0; the transitional probabilities between the vowels across word boundaries were .5; and the transitional probabilities between all adjacent syllables (both within and between words) were .5. The syllable inventory of Language A and Language B overlapped by 50%.

Table 1
Design of two languages in Experiments 1, 3, 4, and 5

$[c_1]\ V_1$ $[c_2]$	$[c_3]\ V_2$ $[c_4]$	$[c_5]\ V_3$ $[c_6]$
$[c_1]\ V_4$ $[c_2]$	$[c_3]\ V_5$ $[c_4]$	$[c_5]\ V_6$ $[c_6]$

Vowel-Frames	Consonant-Fillers
Language A	
_a_u_e	[d_][k_][b_]
_o_i_ae	[p_][g_][t_]
Language B	
_u_ae_i	[b_][p_][g_]
_e_o_a	[t_][d_][k_]

For each language, a continuous stream of nonsense words was created by generating a randomized list. Six blocks, each consisting of a constrained random ordering of one token of each of the 16 words in the language, were concatenated into a text in 10 different random orders. This rendered a list of 2,880 syllables. The stipulations for the randomization were that: (a) the same word never occurred twice in a row, and (b) each word-final syllable could be followed only by either of two specific word-initial syllables.

All word boundaries were removed from the text, which was then read by the MacInTalk speech synthesizer, using the text-to-speech application Speaker, running on a Power Macintosh G3 computer. As the synthesizer was uninformed about word boundaries, it did not produce any acoustic cues to the word boundaries. The synthesizer 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 audiotape from the sound output of the Power Macintosh computer and then recorded again into SoundEdit 16 Version 2. Once recorded in SoundEdit, each syllable was edited to 0.20–0.22 s in length, to ensure that there were no differences in syllable length that could correlate with syllable position in the words.

The speech stream was looped five times to create the familiarization stream of approximately 5 min. It contained no pauses and played at a rate of 284 syllables per minute. It was then recorded from the sound output of the Power Macintosh onto minidisk, for subsequent presentation to the research participants.

A two-alternative forced-choice (2AFC) test was designed to assess participants' ability to learn the patterned relationships, as evidenced by their ability to distinguish words from part-words. Part-words were trisyllabic sequences that spanned a word boundary in the continuous speech stream. Part-words were of two types: (a) a 3-1-2 pattern, consisting of the last syllable of one word and the first two syllables of another word, and (b) a 2-3-1 pattern, consisting of the last two syllables of one word and the first syllable of another word. Each of the test words and part-words was generated separately by the MacInTalk speech synthesizer, in the same way as described for the streams above, except that each was generated in isolation. This produced a falling intonation on the final syllable of each item, making each individual word and part-word sound like it was spoken in isolation. Table 2 shows the words and the part-words that were tested for each language.

Table 2
Test words and test part-words in Experiments 1, 3, 4, and 5

	Words	Part-Words
Language A	da ku be	bae pa gu
	pa gu te	te do ki
	po gi tae	ku be po
	do ki bae	gi tae da
Language B	bu pae gi	ga tu dae
	te do ka	ki be po
	be do ga	pae gi te
	tu pae ki	do ka bu

To design the 2AFC test, four of the 16 words were paired exhaustively with four part-words for each language, to determine whether participants could recognize the more statistically consistent patterns (the words). Each word/part-word combination occurred once during the test, rendering a total of 16 test pairs. In this experiment and in all subsequent experiments, two different randomized presentation orders were used for the test items (counterbalanced across participants). Each test trial paired a word with a part-word and was recorded to minidisk with a 1-s silent interval between each word/part-word test pair and a 5-s silent interval between each test pair.

2.1.3. Procedure

All participants were tested individually in a quiet room. The familiarization streams and test stimuli were presented to participants using headphones connected to a Sony minidisk player. Participants were instructed to listen attentively to a recording of a continuous speech stream that would sound a little like a foreign language. They were told that as they listened, parts of it might become familiar and that after the recording stopped, they would be tested to determine how well they recognized some of the patterns in the recording. After the recording had ended, participants were given a 2AFC test with 16 test trials (as described above). On each test trial, participants heard a word and a part-word. Participants were instructed to indicate which one was more familiar to them, based upon the recording that they had heard, by circling either the ''1'' or ''2'' on a preprinted answer sheet.

2.2. Results and discussion

Participants readily acquired the language to which they were exposed. For each condition (Language A or Language B exposure), performance on the 2AFC test significantly exceeded chance [Language A: $M = 12.75$ out of 16, or 79.69% correct, $t(7) = 4.77$, $p = .002$; Language B: $M = 12.63$ out of 16, or 78.91% correct, $t(7) = 6.13$, $p = .0005$]. There was no statistically significant difference between the performance of participants who had been exposed to Language A and those who had been exposed to Language B [$t(14) = 0.10$, $p = .92$, *ns*]. Fig. 1 shows the pooled mean accuracy level for the participants tested in the Language A alone condition and the participants tested in the Language B alone condition [$M = 12.69$ out of 16, or 79.30% correct, $t(15) = 7.77$, $p < .0001$].

Thus, in the present experiment, we identified two individual languages (Language A and Language B), each of which was learned on its own with 5 min of exposure and with approximately equal accuracy (80% correct). The results demonstrate that participants do not learn one language better than another language when they are exposed to either language alone, and therefore participants should not be predisposed to learn one of the two languages better than the other in subsequent experiments (irrespective of the order of exposure to the languages or to cues marking a switch from one language to another). In subsequent experiments, we exposed a separate set of participants to these two languages successively (with the switch in language occurring in midstream) to determine whether participants could learn both structures under these circumstances.

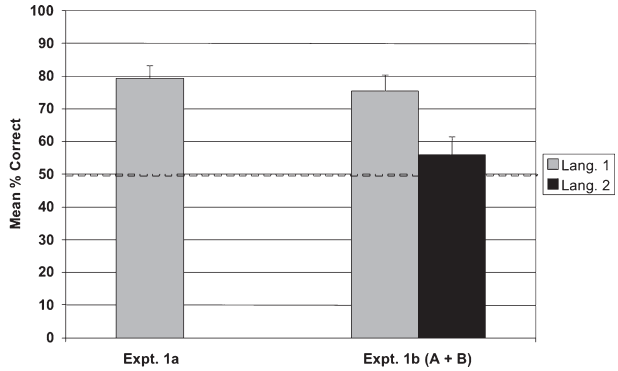

Fig. 1. Group mean accuracy in Experiment 1. For control Experiment 1a, the bar shows group mean accuracy (+*SEM*) on Language 1 alone (this represents the pooled data from the Language A alone and the Language B alone conditions; see text). For Experiment 1b, the bars show group mean accuracy (+*SEM*) on Language 1 and Language 2.

3. Experiment 1b: Language A + Language B

The results of the previous experiment demonstrated that participants were able to learn Language A or Language B with approximately equal accuracy, when they were exposed to each language alone. The purpose of the present experiment was to determine whether a separate set of participants could learn the patterned relationships in both Languages A and B when they were exposed to one language followed immediately by the other language, with no explicit information about when the switch in language occurred or even that there were two structures.

As the languages were both learned when presented individually, the presentation of them in sequence will allow us to ask whether two different languages can be learned, perhaps cued by the differences in their statistical structure, or whether the sequential presentation of the two languages results in a change of learning outcomes. If there is an interaction in the outcome of learning of two successively presented languages, do learners pool the entire exposure corpus, resulting in no learning at all (since the statistics of the entire corpus taken together show no consistency of words over part-words), or rather do learners acquire only the first language or only the second, and at the same level of performance as when each language is presented alone? If the latter, it would suggest that learners use only 5 min or less of an exposure corpus to determine the structure of the language, making their decisions about the underlying structure early (and in this case precipitously).

3.1. Method

3.1.1. Participants

Sixteen University of Rochester undergraduates participated in this experiment for a payment of $10 each. Eight participants were assigned to Condition 1 (5 min Language A

exposure + 5 min Language B exposure), and eight participants were assigned to Condition 2 (5 min Language B exposure + 5 min Language A exposure).

3.1.2. Design and materials

The exposure stream consisted of the 5-min Language A stream (described in Experiment 1a), followed immediately by the 5-min Language B stream (described in Experiment 1a), or vice versa. The Language A and Language B streams were concatenated, with no pause between the streams, using SoundEdit 16 Version 2. This melding of the two language streams was done at a zero-crossing in the waveform file to eliminate any obvious transients, but it did create an absence of coarticulation at the transition from the last syllable of Language A to the first syllable of Language B.[2] As noted in Experiment 1a, the statistical structure in both Language A and Language B resides entirely in the vowel-to-vowel transitional probabilities (1.0 and 1.0 within each word). The vowel-to-vowel transitional probabilities in the tested part-words are .5 and 1.0. Within each language, the consonant-to-consonant and the syllable-to-syllable transitional probabilities carry no information about word boundaries (i.e., they are .5 across each language stream, both within and between words).

Because the design of Experiment 1b involves successive exposure to Language A and Language B, it is important to control the aggregate statistics across both languages. The mean aggregate vowel-to-vowel transitional probabilities in the tested words are .50 and .50 (range: .24–.76) and in the tested part-words the means are 0.44 and 0.44 (range: 0.24–0.76). All syllable-to-syllable transitional probabilities are .50, both within and between words.

The 2AFC test contained 32 trials. As in Experiment 1a, for each language, four words were paired exhaustively with four part-words. In the present experiment, each word/part-word combination occurred once in the 2AFC test, rendering 16 test trials per language, for a total of 32 test trials. The test trials alternated in testing Language A word/part-word discriminations and Language B word/part-word discriminations.

3.1.3. Procedure

The procedure was identical to that in Experiment 1a, except that the familiarization stream actually contained two languages (as described above) and the 2AFC test given at the end of the full 10-min corpus contained 32 test trials (half of which tested Language A word/part-word discriminations and half of which tested Language B word/part-word discriminations). The instructions were identical to the instructions in Experiment 1a.

3.2. Results and discussion

As illustrated in Fig. 1, participants learned the statistical regularities in the first language to which they were exposed but did not learn those of the second language. The pooled performance data from both conditions show that participants learned Language 1 at a level that significantly exceeded chance [M = 12.06 out of 16, or 75.39% correct, $t(15) = 5.16$, $p = .0001$] but did not learn Language 2 [M = 8.94 out of 16, or 55.86% correct, $t(15) = 1.06$, $p = .30$, *ns*]. Participants' performance on Language 1 was significantly better than their performance on Language 2 [$t(15) = 2.70$, $p = .017$]. Moreover, mean

performance on Language 1 when followed by Language 2 (75.39%) did not differ from mean performance in Experiment 1a on Language 1 alone (79.30%), $t(30) = -0.63$, $p = .53$.

These results provide evidence that, in the absence of a robust cue signaling the midstream switch from the first to the second language, the learning of the patterned relationships in the first language blocks the learning of the patterned relationships in the second language. One such cue that could mark the switch from the first to the second language is a change in structure. Recall that the two languages used in Experiment 1 had the same structure (nonadjacent vowel frames with variable intervening consonants), but with different pairings of consonants and vowels that produced a somewhat different inventory of syllables (50% overlap). Perhaps this structural similarity hindered the triggering of the change-detection process. To address this issue, we conducted a second experiment that in all respects was similar to Experiment 1, except that the languages differed significantly in their underlying structures while retaining their surface property of being composed of trisyllabic words. If learners are sensitive to this change in structure, then perhaps they would acquire both structures.

4. Experiment 2a: Language A or Language C alone

We again relied on the Newport and Aslin (2004) studies to select two languages that had different statistical structures. We retained one language (A) from Experiment 1 that was based on a vowel frame and selected a second language (C) whose structure was based on a consonant frame. As shown in Table 3, both languages had a trisyllabic surface structure,

Table 3
Design of two languages in Experiment 2

Language A:	
$[c_1]\ V_1\ [c_2]$ $[c_3]\ V_2\ [c_4]$ $[c_5]\ V_3\ [c_6]$	
$[c_1]\ V_4\ [c_2]$ $[c_3]\ V_5\ [c_4]$ $[c_5]\ V_6\ [c_6]$	
Language C:	
$C_1\ [v_1]\ [v_2]$ $C_2\ [v_3]\ [v_4]$ $C_3\ [v_5]\ [v_6]$	
$C_4\ [v_1]\ [v_2]$ $C_5\ [v_3]\ [v_4]$ $C_6\ [v_5]\ [v_6]$	
Frames	**Fillers**
Language A	
_a_u_e	[d_][k_][b_]
_o_i_ae	[p_][g_][t_]
Language C	
t_d_k_	[_ae][_a][_i]
b_p_g_	[_e][_o][_u]

but they differed in how the same inventory of consonants and vowels was organized to form words. The goal of Experiment 2a was to ensure that these languages were learned to equal levels of proficiency in 5 min of exposure, to parallel the results of Experiment 1a with Languages A and B.

4.1. Method

4.1.1. Participants

Sixteen University of Rochester undergraduates participated in this experiment for a payment of $10 each. Eight participants were assigned to Condition 1 (5 min Language A exposure), and eight participants were assigned to Condition 2 (5 min Language C exposure).

4.1.2. Design and materials

Table 3 illustrates the design and composition of each language. Language A was the same language that had been used in Experiment 1. It consisted of 16 trisyllabic words, constructed from two unique three-vowel frames with two different consonants possible in each of the consonantal positions (refer to ''Design and Materials'' section of Experiment 1a for more details). Language C consisted of 16 trisyllabic words, constructed from two unique consonant-frames with two different vowels possible in each of the vowel positions. This language was also taken from an earlier study (Newport & Aslin, 2004). The transitional probabilities between the consonants within a word were 1.0; the transitional probabilities between the consonants across word boundaries were .50; and the transitional probabilities between syllables (both within and between words) were .50.

For Language C, a continuous stream of words was created by generating a randomized list, using procedures similar to those of Experiment 1a. A stream was formed from six blocks, each of which consisted of a constrained random ordering of one token of each of the words in the language. The six blocks were concatenated into a text in 10 different random orders, which rendered a list of 2,880 syllables. The randomization was performed with the stipulations that the same word never occurred twice in a row, and that each word-final syllable could be followed only by either of two specific word-initial syllables. After recording and editing the stream, it contained no pauses and played at a rate of 280 syllables per minute. The stream was looped five times to render a familiarization stream of approximately 5 min.

As in Experiment 1a, each word was paired with each part-word to create a 16-item 2AFC test (refer to Table 4 for test items).

4.1.3. Procedure

The procedure was identical to that in Experiment 1a, in which a 5-min exposure to only one language was presented and tested, except that in this experiment Language C was substituted for Language B and the test items were as listed in Table 4.

Table 4
Test words and test part-words in Experiment 2

	Words	Part-Words
Language A	da ku be	ku te pa
	pa gu te	ki tae po
	po gi tae	bae do gi
	do ki bae	be da gu
Language C	tae da ku	da ki te
	te do ki	do ku tae
	bae pa gu	gi be pa
	be po gi	gu bae po

4.2. Results and discussion

When only one 5-min language was presented, participants readily acquired the language. For each condition (Language A or Language C exposure), performance on the 2AFC test significantly exceeded chance [Language A: $M = 11.88$ out of 16, or 74.22% correct, $t(7) = 2.62$, $p = .03$; Language C: $M = 13.63$ out of 16, or 85.16% correct, $t(7) = 7.97$, $p < .0001$]. There was no statistically significant difference between the performance of participants who had been exposed to Language A alone and those who had been exposed to Language C alone [$t(14) = -1.07$, $p = .30$, *ns*]. Fig. 2 shows the pooled mean accuracy level for the participants tested in the Language A alone condition and the participants tested in the Language C alone condition [$M = 12.75$ out of 16, or 79.69% correct, $t(15) = 5.76$, $p < .0001$].

Thus, the results of the present experiment paralleled those of Experiment 1a by showing that Languages A and C were each learned on their own within 5 min of exposure and with approximately equal accuracy ($\sim$ 80% correct). The results from this experiment demonstrate that participants should not be predisposed toward learning Language A or Language

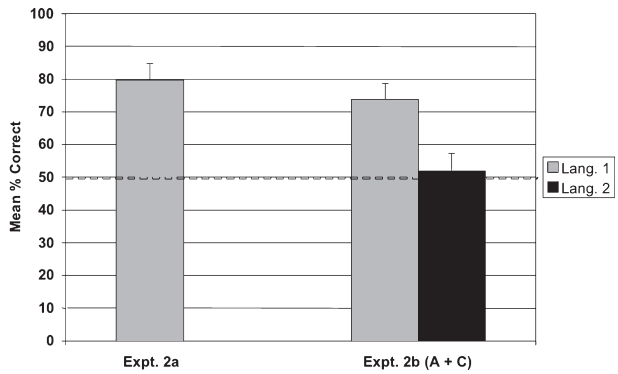

Fig. 2. Group mean accuracy in Experiment 2. For control Experiment 2a, the bar shows group mean accuracy (+*SEM*) on Language 1 alone (this represents the pooled data from the Language A alone and Language C alone conditions; see text). For Experiment 2b, the bars show group mean accuracy (+*SEM*) on Language 1 and Language 2.

C better in Experiment 2b, when the languages were presented successively (irrespective of the order of exposure to the languages). In Experiment 2b, we exposed a separate set of participants to both Language A and Language C (with the switch in language occurring in midstream) to determine whether they could learn both languages.

5. Experiment 2b: Language A + Language C

This experiment investigated whether participants would be able to learn two languages whose statistical regularities had a different structural organization than those that characterized the two languages in Experiment 1b. If participants are sensitive to this qualitative change in the structure of the words, and it automatically triggers the change-detection process, then their performance on the second language should be enhanced relative to that of Experiment 1b. However, if sensitivity to this change in structure enables the learning of the second language, it may have a cost in terms of attenuating performance on the first language. Thus, Experiment 2b not only tests whether a change in language structure triggers a second representation but also how these two representations interact.

5.1. Method

5.1.1. Participants

Sixteen University of Rochester undergraduates participated in this experiment for a payment of $10 each. Eight participants were assigned to Condition 1 (5 min Language A exposure + 5 min Language C exposure), and eight participants were assigned to Condition 2 (5 min Language C exposure + 5 min Language A exposure).

5.1.2. Design and materials

The design and materials for this experiment were identical to those in Experiment 1b, except that Language C was substituted for Language B in the familiarization stream and that different test items were used (see Table 4).

5.1.3. Procedure

The procedure was identical to that for Experiment 1b, except for the substitution of Language C for Language B and for the test items (see Table 4).

5.2. Results and discussion

As illustrated in Fig. 2, participants learned the statistical regularities of the first language to which they were exposed, but they did not learn the second language. The pooled performance from both conditions showed that participants learned Language 1 at a level that significantly exceeded chance [M = 11.81 out of 16, or 73.83% correct, $t(15) = 4.91$, $p = .0002$], but did not learn Language 2 [M = 8.31 out of 16, or 51.95% correct, $t(15) = 0.36$, $p = .72$, *ns*]. Participants' performance on Language 1 was significantly better

than their performance on Language 2 [$t(15) = 2.38$, $p = .03$]. Moreover, there was no statistically significant decline in performance on Language 1 when it was followed by Language 2, compared to performance on Language 1 alone. When the first language was A (vowel-based structure), mean performance declined slightly (but not statistically significantly) from 74.22% to 68.75%, $t(14) = 0.44$, $p = .67$; when the first language was C (consonant-based structure), mean performance declined slightly (but not statistically significantly) from 85.16% to 78.91%, $t(14) = 0.99$, $p = .34$.

These results are similar to those of Experiment 1b in showing that the learning of the first statistical structure (language) blocks the learning of the second statistical structure (language), even when the two languages contain qualitatively different structures. There was no evidence that participants learned the structure of the second language, and an ANOVA across Experiments 1b and 2b revealed no main effect, $F(1,60) = 0.28$, $p = .60$, *ns*, or interaction, $F(1,60) = 0.05$, $p = .82$, *ns*, suggesting that a qualitative shift in structure did not enhance the learning of the second language or attenuate the retention of the first language. Taken together, then, the results of Experiments 1b and 2b provide evidence for a strong primacy effect in the learning of the first language and highlight the resiliency of learning the first structure, as well as its tendency to block the learning of the second structure, despite equal exposure to the two.

A final intriguing aspect of the statistics of Languages A and C concerns the alternative strategy of simply aggregating across both languages in an attempt to use any overall statistical information (at the segment or the syllable level) to learn the two sets of words. In Experiment 1b, such an aggregated statistics strategy would not enable participants to select words over part-words. All relevant statistics—adjacent syllables, nonadjacent syllables, adjacent segments, and nonadjacent segments—failed to differentiate between words and part-words. However, in Experiment 2b, one aggregate statistic computed over Languages A and C did contain information that would allow participants to choose words over part-words. This statistic was the transitional probability of adjacent syllables, which came from exactly the same inventory (100% overlap) between the two languages, but in different sequential orderings. When aggregated over the two 5-min streams, the transitional probabilities between adjacent syllables were .5 within words, .33 between words, and .25 within part-words. To be clear, this adjacent-syllable statistic was *not* informative within either of the two languages (the relevant within-language statistic was nonadjacent segments). Thus, in addition to showing that learners do not detect a structural change and add a second representation when the individually language-relevant statistic undergoes a change at midstream, the results of Experiment 2b also show that participants do not aggregate their statistical computations across the entire combined corpus, even when such a strategy would enable them to learn to discriminate each set of words from their part-words.

6. Experiment 3: Language A + Language B with pause cue

In Experiments 1b and 2b, participants had been instructed that they would be hearing one language (not two languages, as reflected in their actual structures), and there was no pause

between the two languages. In the present experiment, we investigated whether explicitly instructing participants that they would be hearing two languages and also adding a pause to indicate when the transition between Language 1 and Language 2 occurred would help them to learn both languages. Thus, our goal was to determine whether, under unambiguous conditions, participants were capable of learning two successive language structures. In this experiment and those to follow, we focused on Languages A and B because we found no advantage (or disadvantage) to using two languages that shared the same vowel-frame structure.

6.1. Method

6.1.1. Participants

Sixteen University of Rochester undergraduates participated in this experiment for a payment of $10 each. Eight participants were assigned to Condition 1 (5 min Language A exposure + 30-s silent pause + 5 min Language B exposure), and eight participants were assigned to Condition 2 (5 min Language B exposure + 30-s silent pause + 5 min Language A exposure).

6.1.2. Design and materials

The design and materials for this experiment were identical to those for Experiment 1b, except that a 30-s silent pause separated the Language A and Language B streams.

6.1.3. Procedure

The procedure was identical to that in Experiment 1b, except that participants were instructed that they would hear two different languages with a 30-s silent pause between the languages. As in Experiment 1b, participants were asked on each test trial to indicate which of the two combinations of sounds was more familiar to them, based upon the recording that they had heard. They were not required to identify the source (i.e., the first or second language that they had heard) on each test trial.

6.2. Results and discussion

As shown in Fig. 3, participants in this experiment learned both languages. Participants performed significantly better than chance on both Language 1 [M = 11.06 out of 16 or 69.14% correct, $t(15) = 4.28$, $p = .0007$] and Language 2 [M = 11.5 out of 16 or 71.88% correct, $t(15) = 5.37$, $p < .0001$]. The difference between participants' performance on Language 1 versus Language 2 was not statistically significant [$t(15) = -0.45$, $p = .66$, *ns*]. Moreover, mean performance on Language 1 when followed by Language 2 (69.14%) was not statistically less than performance from Experiment 1b on Language 1 alone (75.39%), $t(30) = 0.94$, $p = .36$.

These results illustrate that participants are capable of learning the different structures contained in two successively presented languages when explicit information is provided about the number of languages and the timing of the structural transition. It is not clear whether explicit information about the number of languages, the pause that signaled the

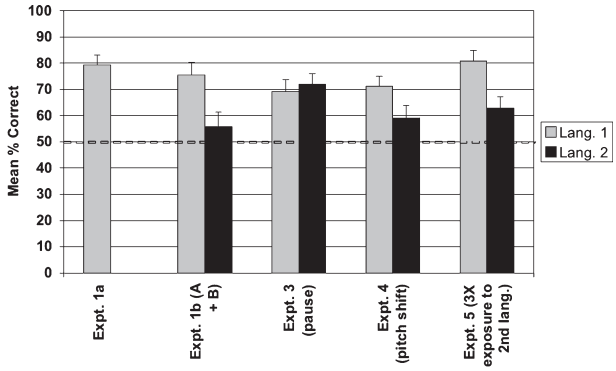

Fig. 3. Group mean accuracy in Experiments 1, 3, 4, and 5. For control Experiment 1a, the bar shows group mean accuracy (+*SEM*) on Language 1 alone (this represents the pooled data from the Language A alone and the Language B alone conditions; see text). For Experiments 1b, 3, 4, and 5, the bars show group mean accuracy (+*SEM*) on Language 1 and Language 2.

change in language, or a combination of these two factors enabled the participants to learn both languages. An interesting question for future study would be to determine which of these contextual cues is necessary for learning two structures.

This led us to ask how the experimental manipulation of other contextual variables would impact learning of successive structures. In the next experiment, we asked whether a different implicit cue could be used, instead of a pause cue plus explicit instructions, to induce the learning of both statistical structures in successive languages.

7. Experiment 4: Language A + Language B with pitch shift cue

In Experiment 3, we found that participants learned both languages when they were explicitly told that there would be two languages and when a pause cue separated the two languages. This raised the question of whether such an explicit cue to the midstream change in structure was necessary for participants to learn both languages, or whether an *implicit* cue could be employed as an alternative to achieve the same result. Experiment 4 tested whether using a noticeable pitch shift cue to mark the transition from Language 1 to Language 2 would enable participants to learn both structures. As in Experiment 1b, participants were told neither that the speech stream contained two languages, nor that the pitch of the voice would shift midstream. However, due to the pitch shift, the voices of Language 1 and Language 2 did sound distinctively different.

7.1. Method

7.1.1. Participants

Sixteen University of Rochester undergraduates participated in this experiment for a payment of $10 each. Eight participants were assigned to Condition 1 (5 min Language A

exposure + 5 min pitch-shifted Language B exposure), and eight participants were assigned to Condition 2 (5 min Language B exposure + 5 min pitch-shifted Language A exposure).

7.1.2. Design and materials

The design and materials for this experiment were identical to those for Experiment 1b, except that the pitch of the synthetic voice for Language 2 was downshifted by 1.1 octaves, relative to the pitch of the synthetic voice for Language 1, using Amadeus II Version 3.8.3 software. The downshifted voice sounded like that of the same talker, but with a deeper tone. The words and part-words from the language that was pitch-shifted in the familiarization stream were also pitch-shifted in the 2AFC test.

7.1.3. Procedure

The procedure was identical to that of Experiment 1b.

7.2. Results and discussion

As shown in Fig. 3, participants learned Language 1 [M = 11.38 out of 16 or 71.09% correct, $t(15) = 5.46$, $p < .0001$] but did not learn Language 2 [M = 9.44 out of 16 or 58.98% correct, $t(15) = 1.83$, $p = .087$, *ns*]. The difference between participants' performance on Language 1 and Language 2 was marginally significant [$t(15) = 2.06$, $p = .058$]. These results demonstrate that an implicit pitch shift cue was insufficient to enable participants to learn both languages, although with additional exposure to Language 2, the marginally significant effect of a pitch change would possibly disappear. A post-hoc power analysis, with N = 16, $\alpha = 0.05_{2\ \text{tail}}$, and effect size dz = 0.48, showed that the power of this experiment in showing a Language 1 versus Language 2 difference was 44.23%. It is possible that with a larger N, the difference between participants' performance on Language 1 and Language 2 would be statistically significant. However, mean performance on Language 1 when followed by a pitch-shifted Language 2 (71.09%) was not statistically different from performance in Experiment 1b on Language 1 (75.39%), $t(30) = 0.69$, $p = .50$, in which no pitch shift had occurred in Language 2.

The design of Experiment 4 was similar in some respects to that employed by Weiss et al. (2009), but a critical difference was that we used a single shift in the underlying structure at midstream rather than repeatedly alternating the structures. In both Experiment 4 and Weiss et al. (2009), the change in structure was cued by a change in voice quality, but we found no evidence of learning the second structure, while Weiss et al. found that both structures were learned. Two design differences are likely relevant in accounting for this difference in learning. First, the structural alternation (in 2-min blocks) used by Weiss et al. allows the learner repeated samples of the voice cue, thereby highlighting the possibility that this contextual cue is correlated with a change in structure. Second, the inventory of sounds that defined the two structures had less overlap in Weiss et al. than the inventory that we used in Experiment 4. Thus, there was a second contextual cue for the change in structure (in addition to voice quality) in Weiss et al. that could trigger a change-detection mechanism at each alternation.

The results from Experiments 1–4 provided clear evidence that the learning of the first language was resilient to the potential interfering effects of the second language, but that, in the absence of a pause cue with explicit instructions, learning the first language appeared to block the learning of the second language. This is evidence that learners formulate implicit hypotheses about the underlying structure of the patterned materials over less than the first 5 min of the exposure corpus (in accord with the results in Experiment 1a from testing each language alone), and also evidence for a strong primacy effect, such that the first 5 min of exposure acts to substantially block or attenuate any learning of the patterns in the second. However, in the experiments thus far, exposure to the two languages was of equal duration. Our next question focused on whether presenting participants with additional countervailing data following the first language (in the form of increased duration of exposure to the second language, relative to the first language) would impact participants' learning of the successive structures.

8. Experiment 5: Language A + Language B with tripled exposure

In the previous two experiments, we identified one cue (a pause plus explicit instructions) that appeared to help participants learn both languages and another cue (pitch shifting the second language) that did *not* appear to help participants to learn both languages. In the present experiment, we investigated how presenting participants with additional countervailing data (by tripling participants' exposure to Language 2 relative to Language 1), while holding all other variables constant, would impact participants' learning of the two successive languages.

If participants learned the first language but not the second language (as in Experiment 1b), this would provide evidence that the learning of Language 1 was relatively impervious to the additional countervailing input (i.e., a primacy effect). By contrast, if participants learned Language 2 but not Language 1, this would suggest that the additional countervailing input helped participants to learn the second structure but also prevented or erased the learning of the first structure (i.e., a recency effect similar to catastrophic interference in connectionist networks). If participants learned both languages, this would provide evidence that the additional countervailing data enables participants to learn a second structure, while retaining the first structure. If, however, participants failed to learn either Language 1 or Language 2, this would suggest that the competition between the learning of the two structures prohibited participants from learning either structure.

8.1. Method

8.1.1. Participants

Sixteen University of Rochester undergraduates participated in this experiment for a payment of $10 each. Eight participants were assigned to Condition 1 (5 min Language A exposure + 15 min Language B exposure), and eight participants were assigned to Condition 2 (5 min Language B exposure + 15 min Language A exposure).

8.1.2. Design and materials

The design and materials were identical to those for Experiment 1b, except that the exposure duration for Language 2 was tripled. Thus, participants were exposed to approximately 5 min of Language 1 and to 15 min of Language 2, with no pause or other cue between the two languages. This is the only exposure duration ratio of Language 1 to Language 2 (i.e., 5 to 15 min) that we tested and was selected to provide participants with the minimum duration of exposure to Language 2 that we thought might result in learning that language, while ensuring that the total exposure duration would not be too long to maintain participants' attention during the experiment. As noted in the design of Experiment 1b with equal durations of exposure to Languages 1 and 2, there were no reliable statistical cues to words over part-words, at either the vowel-to-vowel or the syllable-to-syllable levels, when aggregated across the entire 10-min corpus. In the present experiment, however, there was the potential for the greater exposure to Language 2 (15 min) to introduce an aggregate statistic that could lead learners to perform at above-chance levels on the posttest. However, these differences when the corpora are aggregated are very small: The aggregate vowel-to-vowel transitional probabilities when Language A was followed by Language B were .63 and .50 for words and .50 and .57 for part-words. And when Language B was followed by Language A, these aggregate vowel-to-vowel statistics were .62 and .50 for words and .52 and .62 for part-words. All syllable-to-syllable transitional probabilities for words and part-words were similarly comparable (approximately .50) across the combined Language A + B corpora in the present experiment.

8.1.3. Procedure

The procedure was identical to that in Experiment 1b.

8.2. Results and discussion

As shown in Fig. 3, participants performed above chance on both languages, but they performed much better on Language 1 [M = 12.94 out of 16 or 80.86% correct, $t(15) = 7.54$, $p < .0001$] than on Language 2 [M = 10.06 out of 16 or 62.89% correct, $t(15) = 3.12$, $p = .007$]. The difference between participants' performance on Language 1 versus Language 2 was statistically significant [$t(15) = 2.83$, $p = .013$]. Moreover, mean performance on Language 1 when followed by three times the exposure to Language 2 (80.86%) did not lead to a decline in performance on Language 1 alone (75.39% from Experiment 1b), $t(30) = -0.85$, $p = .40$, in which participants had an equal duration of exposure to Language 1 and Language 2.

The results show that participants learned Language 2 at a level that was statistically above chance, when they had had three times as much exposure to Language 2 than to Language 1. However, participants learned Language 1 much better than they learned Language 2. This suggests that a mere 5 min of exposure to Language 1 serves to partially block the learning of Language 2, even with 15 min of exposure to Language 2. (The above-chance performance on Language 2 might be due to the ratio of Language 1:Language 2 exposure or to the absolute duration of exposure to Language 2.) Thus, there is clear evidence for a

primacy effect for the learning of Language 1 but also for a recovery effect in the learning of Language 2 (compared to Experiment 1b) that did not interfere with the structures learned from Language 1. An interesting question for future work is whether both languages would be learned at a level commensurate with that in Experiment 1a, if participants were exposed to Language 2 for a much longer duration or to a Language 2:Language 1 ratio greater than 3:1.

9. General discussion

In a series of five experiments we showed that human learners have a strong bias to acquire and retain the first of two successive statistical structures when there is no explicit cue to signal a midstream shift in these structures. These results suggest that despite significant structural cues and surface cues (see note 2), involving a shift in how consonants and vowels are organized to form syllables and words, and even when these structural cues are accompanied by a change in pitch, learners do not spontaneously form a second structural representation or model into which the postshift input corpus is directed and separately learned. Importantly, however, when the midstream shift in structure is made explicit by informing the learner that there are two languages and by adding a short pause at midstream to make the structural transition, learning of both structures is obtained and there is no loss of performance for either of the two languages compared to a single-language baseline. This demonstrates not only that learners have the capacity to learn and retain two successive structures when they have been explicitly labeled but also that interference between the two structures under the exposure and testing conditions employed in our studies is minimal and cannot be the primary explanation for failure to learn the second language.

Our final experiment demonstrated that learners can show sensitivity to countervailing evidence that a second structure is present that differs from the first structure, if the evidence is strong or persistent enough. When exposure to the second language was tripled in duration compared to the first language, performance on the second language rose above chance. Importantly, performance on the first language was still maintained at a level indistinguishable from performance on the first language alone. These results show not only that learners detect a structural change, but that they create two separate structural representations for the two subsets of the corpus. Moreover, as in the case of an explicit cue at midstream shift, the slow emergence of a second structural representation does not result in interference with the representation retained from exposure to the first structure. Importantly, in none of the experiments reported here using streams of speech was there any significant attenuation of learning the first structure, even when, as in Experiment 5, extensive exposure to the second structure rose to above-chance levels.

This pattern of results bears on the key issue raised in the Introduction to this article: what is the trade-off between *inefficiency* in waiting for a substantial corpus of input from which structure is extracted and *error recovery* if initial learning is biased by the sparse sampling of the first input that learners happen to receive (i.e., the statistical garden path)? As expected, learners do not wait for an input corpus to ‘‘settle’’ into a uniform underlying

distribution, but rather they quickly acquire structures from limited subsamples of a much larger corpus. Unfortunately, learners require a much longer exposure to the second of two successive structures to begin the error-recovery process. Thus, statistical learning of speech streams places more weight on efficiency than on error recovery, thereby exhibiting a garden-path effect. Importantly, this garden-path effect is not inevitable: Learners are capable, under some conditions, of learning both of the successively presented structures (i.e., when given explicit cues). Thus, two separate structural representations can be triggered by some contextual cues.

9.1. What triggers structural change detection?

The results of Experiment 3 demonstrate that adult learners have the capacity to acquire two successive structures when the midstream change is signaled by explicit instructions and a short pause. However, the more natural case is when a structure undergoes a change implicitly, or when there is high variance around a single structure. When two successive structures were of equal duration, as in Experiments 1b and 2b, a second structural representation was not triggered. Similarly, the midstream pitch cue introduced in Experiment 4 was not effective in triggering a second structural representation. This might have been because adults in a monolingual environment do not expect differences between talkers or differences in tone within a single talker to signal a structural change. While naïve learners (e.g., infants) cannot know a priori that talker differences are typically irrelevant to a change in structure, they apparently learn this quite early. In a discrimination task, 6-month-olds weight talker differences as less important than changes in phonetic (vowel) category (Kuhl, 1979), and in a word recognition task the performance of 7.5-month-olds is adversely affected by talker variation, but the performance of 10-month-olds is not (Houston & Jusczyk, 2000). These results suggest that by the end of the first year of life, infants have learned that talker differences typically do not signal changes in language-relevant structure.

As noted above, the results of Experiments 1 and 2 provide clear evidence that a midstream change in structure is insufficient to trigger a second structural representation. There were two sources of information about this structural change. The first source of information is the inventory of syllables, which in Experiment 1 had only a 50% overlap between the two successive languages. The second source of information is the statistical structure by which words were defined in the two successive languages. In Experiment 2 this structure was based on either vowel frames (as in Experiment 1 for both languages) or consonant frames. Why did neither of these structural triggers induce the formation of a second structural representation at midstream? One possibility, as in the case of a change in pitch, is that the phonetic inventory, despite only 50% overlap, was entirely consistent with English. Thus, monolingual adults may have treated the midstream shift as merely a different subset of the larger corpus of English phonemes. If the phoneme inventory had undergone a midstream change, then perhaps that would trigger a second structural representation. This is the more typical bilingual input situation, where talkers use a different phonetic inventory and/or a different distribution of phonetic combinations (i.e., phonotactics) when speaking in different languages.

Importantly, the results of Experiment 5 demonstrate that these structural cues, which were insufficient to trigger the formation of a second structural representation when both successive structures were of equal duration, did eventually trigger a second structural representation when the second language was increased in duration by a factor of three compared to the first language. Thus, some implicit error-monitoring mechanism must be at work to evaluate the ''fit'' between the current sample of input and the structure that has already been learned and stored in the first structural representation. Even in the absence of low-level cues that could serve to trigger the formation of a second structural representation, the growing evidence for a mismatch between the first and the second structures is apparently sufficient to establish two structural representations.

9.2. Why maintain the first structural representation?

The overall bias to learn the first, and not the second, of two successive structures suggests that adult learners have a prior probability, either innately or via early experience, that structures do not undergo rapid change without a strong contextual cue. This bias to expect slow changes in structure is analogous to a variety of smoothness constraints in domains such as visual surface perception and the perception of prosody in spoken language. The success of the unsupervised mixture of experts model developed by Kehagias and Petridis (1997) depended crucially on this slow-change bias. A strong contextual cue, such as the presence of an edge in vision or the end of an utterance in spoken language, can serve to violate the smoothness constraint. But there are dozens of potential contextual cues in most natural scenes or spoken utterances, and if all contextual cues triggered a new structural representation, there would be too many representations for a system with finite capacity to utilize them all effectively.

If there is a capacity limit on the number of structural representations, then why retain the first one (primacy) rather than favoring the last one (recency)? One possibility is that if early representations were deleted in favor of later ones, then relearning would be more difficult, and in addition, the first environment that learners encounter might well be the most important. Holding onto an early representation is a hedge in case that former structure becomes relevant at some later time. Thus, it appears that an equal weighting of brief exposure to the first structure and much longer exposure to the second structure is the best compromise between inefficiency and error recovery. Two examples of this retention of an early structural representation that initially appears to serve no useful purpose once the second structural representation has been created come from research on the barn owl and on second language learning by adults.

In a classic series of studies leading up to Knudsen (1998), Knudsen and his colleagues had shown that, during an early sensitive period, the juvenile barn owl acquires the relationship between sound cues and visual cues to the location of an object. During this sensitive period barn owls can also adapt to a shift in this sound–vision relationship induced by an ear plug or ocular prisms. Knudsen (1998) demonstrated that if the animal experiences the shift only during the first half of the sensitive period and is then allowed to recover to the initial state during the second half, then when the former (perturbed) relationship is reintroduced

in adulthood, it is adapted to rapidly and effectively. This supports the notion that the early perturbed experience in the first half of the sensitive period established a ''structural representation'' that was then suppressed by the return to the unperturbed experience in the second half of the sensitive period. Thus, a latent structure established in the first structural representation was retained and used to adapt to the perturbation later in life, thereby providing the organism with a highly adaptive mechanism for later learning.

Another example of an early structural representation that is retained for later use comes from the work of Au, Knightly, Jun, and Oh (2002). They tested a subset of undergraduates who were taking a Spanish course for the first time and who had overheard (but not used) Spanish as an ambient language during infancy. Although most of these undergraduates had never spoken any Spanish, even as toddlers, they had a much easier time acquiring the phonetic categories of Spanish consonants as young adults than most other monolingual speakers of English. Apparently, early exposure to Spanish led them to retain a separate structural representation for the phonetic categories of Spanish consonants. This second structural representation remained latent (they reported virtually no use of Spanish in the intervening 17–19 years), analogous to the perturbed relationship between visual and auditory cues in Knudsen's barn owls. Both of these examples, therefore, illustrate the advantage of retaining the first of two structural representations into which different structures are segregated for learning. Failure to retain this early structural representation would lead to much greater difficulty, and in some cases the inability, to reacquire the initial structure. Of course, these developmental examples may not apply as well to learning in adults that occurs over much shorter time periods. The results of the present series of experiments with adults, therefore, may involve a mechanism for forming separate structural representations that is qualitatively different from the one that operates in early development but achieves a similar outcome.

9.3. Are the structural representations independent or interactive?

Throughout this article we have characterized the representations into which structural information about streams of speech are represented as if, once formed, they are largely independent from each other, with little ''cross-talk'' or interference. A key finding from our experiments supports this notion of independent representations: Learners showed no significant decrement in performance on the first language regardless of performance on the second language (e.g., even when Language 2 was learned in Experiments 3 and 5). However, this finding does not provide conclusive evidence for representational independence. It is possible that learners shift their criterion for what is a relevant level of structure (i.e., from vowels to syllables), even though syllable-to-syllable transitional probabilities have been rendered uninformative by our controlled design.[3] For example, during the first 5 min of exposure in Experiment 1b, learners acquire the vowel-to-vowel structure and retain it through the subsequent 5 min of exposure to a different vowel-to-vowel structure (note again that no syllable-to-syllable structure is present and no aggregate vowel-to-vowel or syllable-to-syllable statistic is useful for discriminating words from part-words). However, if learners are computing syllable-to-syllable level statistics, despite their noninformativeness,

then perhaps the overlap in these syllable statistics between the two languages delays the formation of a second representation.

One test of this hypothesis is to examine how the overlap in syllables between the two languages influences performance on particular test trials. In Experiment 1b, there were two types of test trials: those that involved words and part-words from Language A and those that involved words and part-words from Language B. Because 50% of the syllables in Languages A and B were identical, each test trial consisted of two three-syllable items with some syllables that came from both languages. Among the six syllables in these two test items, the number of syllables that came from both languages ranged from 1–5. Thus, if interference in learning the two languages could be attributed, in part, to shared syllable inventories, one would expect a significant positive correlation between the amount of syllable overlap on subsets of test items and the mean proportion of incorrect choices in the 2AFC word versus part-word test. Across the 16 participants in Experiment 1b, the Pearson's r for syllable overlap (1–5) and proportion incorrect (.0–1.0) ranged from .00 to .654, with a mean of .251. This mean correlation was significantly different from zero, $t(15) = 5.23$, $p < .0001$. Although syllable overlap of test items might have contributed slightly to participants' poor performance on the learning of the second language (the correlation accounts for only about 6% of the variance in posttest performance), it was clearly not a major factor.

The sensitivity of learners to different inventories of elements, whether applied to the same structure or to different structures, is a topic of considerable interest in the artificial grammar learning literature. Studies of 7- and 12-month-old infants by Marcus, Vijayan, Bandi Rao, and Vishton (1999) and Gomez and Gerken (1999), respectively, suggest that transfer of a given grammar from one inventory to another is readily accomplished early in development. Studies of adults suggest that a change in inventory of elements can sometimes signal a change in the grammar (Conway & Christiansen, 2006) and sometimes signal that the grammar is invariant (Lany et al., 2007). Thus, future work should explore the conditions under which contextual cues do and do not support the presence of different structures.

9.4. Computational models of dual structures

In the Introduction, we briefly summarized two computational models that can detect a structural change (Bayesian Blocks and Mixture of Experts). The empirical findings from the present series of experiments suggest that detecting a change in structure is only the first step in partitioning the structures into separate representations. For example, connectionist models provide a poor fit to structures that change across time (Bengio, Simard, & Frasconi, 1994), unless the two input types are labeled. Moreover, most connectionist models do not show primacy effects like those we observed here by human learners, but rather weight the most recent inputs and induce catastrophic interference with early inputs (McCloskey & Cohen, 1989).

Generative models in the Bayesian tradition are good candidates for describing the empirical results from our experiments. The class of models called *particle filters* sample the input on-line (rather than in batch mode) and use the unfolding variance to determine when, and

to which structural model, the input should be assigned (Doucet, Briers, & Senecal, 2006; Doucet, de Freitas, & Gordon, 2001). The key point is to solve this credit-assignment problem—what portion of the variance observed in the input should be attributed to (a) random fluctuations in a single structure, versus (b) two or more structures with separate and reduced variances? Obvious contextual cues that partition the input can serve to trigger more than a single structural representation. But apparently such obvious contextual cues are not necessary, as the lack of fit with a single-structure model can itself trigger a dual-structure model. The precise mechanism by which a second structural representation is triggered, and the initial structural representation is maintained, is a topic of considerable interest for our future research.

9.5. Summary and conclusions

In summary, we have provided clear evidence for a primacy effect in statistical learning that blocks or attenuates the acquisition of a second structure unless the structural shift is marked by an obvious contextual cue. We have also shown that learning a second structure is possible without a midstream cue if the second structure is presented for an extended duration vis-à-vis the duration of the first structure. Finally, we have documented that learning of the second structure had no significant effect (either blocking or attenuation) on the maintenance of the first structure. A question that remains for future investigations is whether further exposure to the second structure would eventually result in the loss of the first structure, or whether (as in the examples from Au et al., 2002 and Knudsen, 1998) the representation of the first structure is immune to subsequent interference. If there is loss of the first structure, it would be interesting to determine whether that structure could be reactivated by minimal re-exposure to it, as in studies of infant memory (Rovee-Collier, 1999).

Notes

1. Subsequent to these studies, we learned that J.A. Catena, B.J. Scholl, P.J. Isola, and N.B. Turk-Browne (personal communication) have conducted a similar experiment in the visual domain. Their stimuli consisted of a sequence of visual shapes, presented one at a time according to a set of temporal-order constraints (see Fiser & Aslin, 2002, for a description of the materials and procedure used by Catena et al.). In contrast to the present series of studies in which there were two structured input sets, in Catena et al. one of the input sets had no structure (i.e., it consisted of a random sequence of shapes). When the structured sequence was followed by the random sequence, performance on a test of whether the structured sequential information was learned was well above chance. However, when the structured sequence was preceded by the random sequence, no evidence of learning the structured information was obtained. In the present series of experiments, we test learning from two equally structured sources and use a more subtle discrimination than the one used by Catena et al. They presented

strings of shapes that were either completely coherent or completely incoherent (never occurred in the input set). In our test, however, we contrast strings of syllables that are completely coherent words with those that are partially coherent part-words that did occur in the input set, thereby demanding a more refined extraction of statistical information about the underlying structure. Moreover, we ask whether learners can perform this discrimination on each of the two structured languages, permitting us to assess the degree to which each of two structured inputs is learned separately, merged with the other, or not learned at all.

2. The absent coarticulation cue as well as the 50% change in syllable inventory at the transition from Language A to Language B (or vice versa) could have enabled learners to detect the presence of the structural transition. To determine whether the languages were different enough to be discriminated from one another on the surface (rather than the statistical) level, we administered six naïve adults (who had had no prior familiarization with the languages) a 32-trial test of their ability to judge whether brief streams of the languages that we had used in the present experiments were the same or different. In each trial, participants heard a 10-s stream of one language, a 1-s pause, and a 10-s stream of a second language. Half of the trials contained identical streams of the same language (e.g., 10 s of Language A and 10 s of Language A); half of the trials contained streams of different languages (e.g., 10 s of Language A and 10 s of Language B). We tested participants' ability to judge whether the following pairs of language streams were the same or different: A-A, B-B, C-C, A-B, and A-C. For trials in which the streams were from different languages, the languages were presented in counterbalanced order (e.g., for the A-B pairing, participants heard Language A first 50% of the time, and Language B first 50% of the time). Participants scored a mean of 85.70% correct overall on the test [$t(5) = 8.27$, $p = .0004$]. Participants' mean performance on each category was above chance: A-A = 87.50% correct ''same'' responses [$t(5) = 4.74$; $p = .0051$]; B-B = 83.33% correct ''same'' responses [$t(5) = 4.00$; $p = .010$]; C-C = 73.62% correct ''same'' responses [$t(5) = 3.58$; $p = .016$]; A-B = 95.83% correct ''different'' responses [$t(5) = 17.39$; $p < .0001$]; A-C = 80.95% correct ''different'' responses [$t(5) = 2.69$; $p = .043$]. This provides evidence that the languages were different enough from one another to be discriminated on the basis of surface-level properties.
3. We thank an anonymous reviewer for this suggestion.

Acknowledgments

This research was supported by NIH grants (HD-37082 and DC-00167) and by a grant from the Packard Foundation to the second and third authors. The first author was supported by a training grant from NIH (T32-MH19942). We thank Joanne Esse and Amanda Robinson for assistance in constructing the speech streams, David Shanks for helpful comments on a draft manuscript, and Nick Chater, Peter Dayan, and Robert Jacobs for clarifying discussions about these findings.

References

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, *91*, 112–149.

Aslin, R. N., & Newport, E. L. (2008). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research directions*. Mahwah, NJ: Erlbaum.

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

Au, T. K., Knightly, L. M., Jun, S. A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, *13*, 238–243.

Barnes, J., & Underwood, B. (1959). ''Fate'' of first-learned associations in transfer theory. *Journal of Experimental Psychology*, *58*, 97–105.

Basseville, M., & Nikiforov, I. (1993). *Detection of abrupt changes – Theory and application*. Englewood Cliffs, NJ: Prentice-Hall.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*, 157–166.

Bialek, W., Nemenman, I., & Tishby, N. (2001). Predictability, complexity and learning. *Neural Computation*, *13*, 2409–2463.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities. *Psychological Science*, *17*, 905–912.

Doucet, A., Briers, M., & Senecal, S. (2006). Efficient block sampling strategies for sequential Monte Carlo. *Journal of Computational and Graphical Statistics*, *15*, 693–711.

Doucet, A., de Freitas, N., & Gordon, N. J. (2001). *Sequential Monte Carlo methods in practice*. New York: Springer-Verlag.

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.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135.

Gustafsson, F. (2000). *Adaptive filtering and change detection*. New York: Wiley.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.

Houston, D. M., & Jusczyk, P.W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception & Performance*, *26*, 1570–1582.

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

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, *3*, 79–87.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Junge, J. A., Scholl, B. J., & Chun, M. M. (2007). How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. *Visual Cognition*, *15*, 1–11.

Kareev, Y., & Fiedler, K. (2006). Nonproportional sampling and the amplification of correlations. *Psychological Science*, *17*, 715–720.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, *126*, 278–287.

Kehagias, A., & Petridis, V. (1997). Time-series segmentation using predictive modular neural networks. *Neural Computation*, *9*, 1691–1709.

Knudsen, E. I. (1998). Capacity for plasticity in the adult owl auditory system expanded by juvenile experience. *Science*, *279*, 1531–1533.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, *66*, 1668–1679.

Lany, J., Gomez, R. L., & Gerken, L. (2007). The role of prior experience in language acquisition. *Cognitive Science*, *31*, 481–507.

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

Marsh, J. K., & Ahn, W. (2006). Order effects in contingency learning: The role of task complexity. *Memory & Cognition*, *34*, 568–576.

McCloskey, M., & Cohen, N. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation, Volume 24* (pp. 109–164). New York: Academic Press.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*, 363–386.

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

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233–238.

Rovee-Collier, C. (1999). The development of infant memory. *Current Directions in Psychological Science*, *8*, 80–85.

Saffran, J. R. (2003). Musical learning and language development. *Annals of the New York Academy of Sciences*, *999,* 397–401.

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.

Scargle, J. D., Norris, J., & Jackson, B. (2006). Studies in astronomical time series analysis. VI. Optimal segmentation: Blocks, triggers, and histograms. Available at: http://trotsky.arc.nasa.gov/~jeffrey/. Accessed September 30, 2008.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology*, *37B*, 1–21.

Shukla, M., Nespor, M., & Mehler, J. (2007). The interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*, 1–32.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*, 303–333.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105–110.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Weiss, D. J., Gerfen, C., & Mitchel, A. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development*, *5*, 30–49.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental model of speech processing. *Language Learning and Development*, *1*, 197–234.