Learnable vs. Unlearnable Harmony Patterns

Regine Lai

The present study provides empirical evidence for Heinz's (2010) Subregular Hypothesis, which predicts that some gaps found in the typology of phonotactic patterns are due to learnability—more specifically, that only phonotactic patterns with specific computational properties are humanly learnable. The study compares the learnability of two long-distance harmony patterns that differ typologically (attested vs. unattested) and computationally (Strictly Piecewise vs. Locally Testable) using the artificial-language-learning paradigm. The results reveal a general bias toward learning the attested, Strictly Piecewise pattern, exactly as the Subregular Hypothesis predicts.

Keywords: phonotactics, learnability, computational phonology, formal theory, typology, dependencies

1 Introduction

This article presents evidence from artificial-language-learning experiments that phonotactic patterns with particular computational properties are more easily learned than patterns without them. The *Subregular Hypothesis* (Heinz 2010) states that only phonotactic patterns belonging to certain Subregular classes are learnable. As explained below, this hypothesis makes strong claims about the nature of possible phonotactic constraints and therefore has implications for any theory of phonology that affirms phonotactic generalizations, including Optimality Theory (OT; Prince and Smolensky 2004) (see section 2.2) and maximum entropy models (Goldwater and Johnson 2003, Hayes and Wilson 2008). If this hypothesis is correct, it helps us better understand what constitutes a possible phonotactic pattern as well as how such patterns can be learned.

Not all logically possible phonological patterns are attested in natural language phonologies. Proposals based on *analytic bias* and *channel bias* aim to account for these gaps. Analytic bias suggests that there are cognitive biases that facilitate the learning of some patterns but suppress the learning of others (Wilson 2003, Moreton 2008). Channel bias, on the other hand, explains the absence of certain patterns as systematic errors that speakers and listeners make, which cause a loss in intended information during transmission (Ohala 1993, Hale and Reiss 2000, Barnes 2002, Blevins 2004). This article does not speak directly to theories based on channel bias or its predictions; it only explores the *dimensions* of analytic bias.

I would like to thank Jeffrey Heinz, Arild Hestvik, William Idsardi, Irene Vogel, members of the University of Delaware Phonology Group, participants at the Linguistic Society of America's 2012 annual meeting, and two anonymous *LI* reviewers for their invaluable comments and suggestions. This article is supported by the NSF DDRIG # 1123610.

Linguistic Inquiry, Volume 46, Number 3, Summer 2015
425–451

doi: 10.1162/ling_a_00188

Moreton (2008) considers the predispositions of Universal Grammar (UG) to be examples of analytic biases (also see Nowak, Komarova, and Niyogi 2002). His research suggests that analytic or inductive bias is strong enough to create typological asymmetries when the channel bias is controlled for. The dimension of analytic bias that Moreton examined was feature-based: he presents evidence showing that a dependency based on the same feature is more readily learned than a dependency based on different features.

This article addresses the existence of another set of cognitive biases: those concerning the *type* of phonotactic constraint, where a constraint's type is determined by its inherent computational properties. For example, a constraint that uses variables for enforcing identity can be thought of as a different type of constraint from one that does not (Berent et al. 2012).

This article reports the results of artificial-language-learning experiments that tested and compared the learnability of two patterns that belong to distinct computational classes, but are otherwise matched in important respects. The results show a clear difference in how experimental participants internalize these two patterns, in accordance with the predictions of the Subregular Hypothesis.

2 The Subregular Hypothesis

2.1 Motivation

The attested typological variation in phonological patterns almost certainly underdetermines the possibilities. What is a possible phonological pattern? Computational analyses of phonological patterns have made significant claims regarding what constitutes a possible phonological pattern. For example, Kaplan and Kay (1994) argue that virtually all phonological patterns belong to the *Regular* class within the Chomsky hierarchy. Regular languages are ones describable by finite state automata.

However, Kaplan and Kay's work does not imply that *all* Regular patterns are possible phonological ones. Heinz (2010) argues that phonotactic patterns actually belong to specific proper subsets of the Regular languages (i.e., Subregular classes), namely, the *Strictly Local* (SL),[1] *Strictly Piecewise* (SP), and *Tier-Based Strictly Local* (TSL) classes (McNaughton and Papert 1971, Heinz 2010, 2011a,b, Rogers et al. 2010, Heinz, Rawal, and Tanner 2011, Rogers and Pullum 2011). Informally, Strictly Local patterns are local dependency patterns, Strictly Piecewise patterns are long-distance dependencies, and Tier-Based Strictly Local patterns are essentially local dependency patterns operating over abstract phonological tiers (these are more formally defined in section 2.2). Figure 1 provides a schematized representation of these kinds of constraints and classifies *Sibilant Harmony* (SH), which is an attested long-distance dependency pattern; *Nasal Place Assimilation*, which is an attested local dependency pattern; and *First-Last Assimilation* (FL), which is an unattested, non–Strictly Local, non–Strictly Piecewise, and non–Tier-

[1] The definitions of Subregular classes and the patterns tested are also given in appendix C.

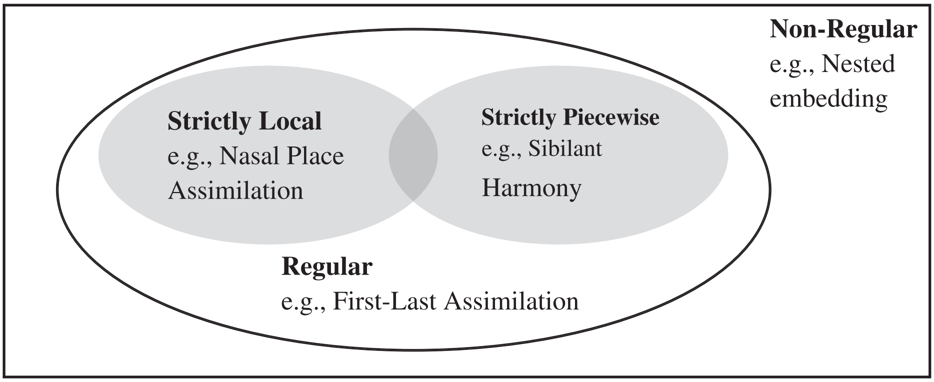

Figure 1
Subregular boundaries. Tier-Based Strictly Local patterns are a proper superset of Strictly Local patterns but a proper subset of Regular patterns, and whether they properly include the Strictly Piecewise patterns is unknown.

Based Strictly Local pattern. First-Last Assimilation roughly corresponds to a long-distance harmony pattern that only requires the first and last sounds of a word to assimilate (for details, see section 2.3). Although terms like *harmony* and *assimilation* usually indicate alternations, these terms are used throughout to refer to the valid phonotactic (surface) generalizations resulting from such alternations.

Why is First-Last Assimilation unattested? If the learning mechanism for phonology can only learn phonotactic constraints that are Strictly Local, Strictly Piecewise, or Tier-Based Strictly Local, as Heinz (2010) suggests, then the absence of patterns such as First-Last Assimilation from the attested languages can be explained: the regularities present in patterns of this type cannot be extracted by humans' phonological learning mechanism. As explained below, the specific patterns tested are well-understood and well-motivated from the perspective of theoretical linguistics and theoretical computer science. The next sections develop these ideas in more depth.

2.2 Strictly Local, Strictly Piecewise, and Tier-Based Strictly Local Patterns

Strictly Local patterns are those that can be described in terms of a finite set of forbidden (contiguous) sequences of symbols of length k (thus, this pattern is called Strictly k-Local). The set of forbidden contiguous sequences can be interpreted as OT-style markedness constraints such as *xy.

On the other hand, Strictly Piecewise languages make distinctions on the basis of (potentially discontiguous) *subsequences* of length k (Heinz 2010, Rogers et al. 2010). A string is a subsequence of another string if and only if its symbols occur in the other string in order. For example, both [ʃʃ] and [oa] are subsequences of [ʃokiʃaʃ], but [ao] is not. Strictly 2-Piecewise languages are those describable by grammars that are sets of forbidden subsequences of length 2. As illustration of a Strictly 2-Piecewise language, consider Sibilant Harmony. Sibilant Harmony requires all sibilants within a word to agree in anteriority; therefore, words obeying this pattern do not contain subsequences of two disagreeing sibilants (i.e., [sʃ] and [ʃs] are forbidden). The set of forbidden potentially discontiguous sequences can be interpreted as OT-style markedness constraints such as *x . . . y.

Tier-Based Strictly Local patterns are essentially Strictly Local ones that operate on an abstract tier projected from the segmental tier. Sibilant Harmony is also a Tier-Based Strictly Local pattern because it can be described as forbidding agreeing contiguous sequences on a sibilant tier. A pattern that is Tier-Based Strictly Local but neither Strictly Local nor Strictly Piecewise is a long-distance disharmony pattern with blocking (Heinz 2010). (See Heinz 2010 for a more detailed discussion on Strictly Local, Strictly Piecewise, and Tier-Based Strictly Local patterns and Rogers et al. 2010 and Heinz, Rawal, and Tanner 2011 for mathematical details.) Tier-based approaches have been employed to solve the problem of learning long-distance phonotactic patterns (Hayes and Wilson 2008, Goldsmith and Riggle 2012).

An example of a logically possible non–Strictly Local, non–Strictly Piecewise, non–Tier-Based Strictly Local but Regular pattern would be one that requires every word to have an even number of sibilants (i.e., words with an odd number of sibilants are disallowed). This pattern cannot be described by a finite set of forbidden sequences or subsequences, not even with any type of phonological tier projection. It follows that phonological learning models that can only learn phonotactic constraints that are Strictly Local, Strictly Piecewise, or Tier-Based Strictly Local will fail to learn this pattern. For example, even given a corpus that robustly exhibits this pattern, the learning model in Hayes and Wilson 2008 will fail to discover it. Whether or not this is a defect of the learning model depends only on whether one thinks it is desirable for phonotactic learning models to discover this type of pattern. Given its implausibility as a humanly possible phonotactic pattern, it seems reasonable not to expect learning models to discover it.

Another, less bizarre Regular pattern that is non–Strictly Local, non–Strictly Piecewise, and non–Tier-Based Strictly Local is First-Last Assimilation. Words obeying this pattern require the first and last sound segments of a word to agree in some feature. As explained further below, this pattern is phonologically plausible. Therefore, it plays a central role in this study.

Henceforth, in this article the term *Subregular* will be reserved specifically to mean patterns belonging to the Strictly Local, Strictly Piecewise, and Tier-Based Strictly Local classes.

2.3 First-Last Assimilation and Sibilant Harmony Patterns

One example of a Regular sound pattern that is not found in any natural language is long-distance assimilation between only the first and last sounds of a word. Unlike the well-documented long-distance harmony patterns (Hansson 2001, Rose and Walker 2004), First-Last Assimilation allows disharmonic intervening segments so long as the first and last sounds are harmonic.

The comparison with Sibilant Harmony, which is documented in Navajo (Sapir and Hoijer 1967), is instructive. Navajo requires sibilants in well-formed words to agree in anteriority. Hypothetical words such as [sototos] and [ʃototoʃ] are both grammatical because the two sibilants in each word agree in anteriority, but *[ʃototos] and *[sototoʃ] are ill-formed because in each case the two sibilants disagree in anteriority. In terms of OT-style markedness constraints, the set of constraints that outputs a Sibilant Harmony language includes *s . . . ʃ and *ʃ . . . s. By contrast, First-Last Assimilation permits both [sototos] and [ʃototoʃ], because the sibilants in the *first* and *last* positions agree in anteriority. However, *[ʃototos] and *[sototoʃ] do not meet this requirement, so they are ill-formed. As the positions of the sibilants affect the grammaticality of a word in First-Last Assimilation, the markedness constraint for this pattern must include boundary symbols: *#s . . . ʃ# and *#ʃ . . . s#.

The difference between Sibilant Harmony and First-Last Assimilation becomes more apparent when examples with sibilants in word-medial positions are examined. First-Last Assimilation predicts that [soʃotos] is well-formed because the first and last sibilants are harmonic. According to Sibilant Harmony, on the other hand, [soʃotos] is ill-formed because the word-medial sibilant disagrees with the others. Table 1 summarizes these examples. Note that all words that are well-formed according to Sibilant Harmony are also well-formed according to First-Last Assimilation (i.e., Sibilant Harmony–acceptable words are a proper subset of First-Last Assimilation–acceptable words).

Computational analysis of these patterns reveals that Sibilant Harmony is Strictly Local (Heinz 2010), but First-Last Assimilation is neither Strictly Local, Strictly Piecewise, nor Tier-Based Strictly Local. First-Last Assimilation belongs to the Locally Testable class of the Subregular hierarchy. The Locally Testable class is a superset of the Strictly Local class and a proper subset of the Regular languages. (For more information on the Subregular hierarchy, see Rogers and Pullum 2011 and Rogers et al. 2013.)

The learnability of First-Last Assimilation can be assessed by comparing it with the learnability of Sibilant Harmony. Sibilant Harmony, which is an attested pattern, differs only minimally

Table 1
Examples of legal and illegal strings according to First-Last Assimilation (FL) and Sibilant Harmony (SH) grammars. Ellipsis is used to show that the sound segments are not necessarily adjacent to each other.

		Strings
FL/SH	Well-formed according to both First-Last Assimilation and Sibilant Harmony	[s . . . s . . . s], [ʃ . . . ʃ . . . ʃ]
FL/*SH	Well-formed according to First-Last Assimilation but ill-formed according to Sibilant Harmony	[s . . . ʃ . . . s], [ʃ . . . s . . . ʃ]
*FL/*SH	Ill-formed according to both First-Last Assimilation and Sibilant Harmony	[ʃ . . . ʃ . . . s], [s . . . ʃ . . . ʃ]
*FL/SH	Ill-formed according to First-Last Assimilation but well-formed according to Sibilant Harmony	None

Table 2
The environments of sibilant cooccurrence in Sibilant Harmony and First-Last Assimilation

	[s]	[ʃ]
[s]	✓	×
[ʃ]	×	✓

Sibilant Harmony: [__ . . . __].

	[s]	[ʃ]
[s]	✓	×
[ʃ]	×	✓

First-Last Assimilation: [#__ . . . __#]

from First-Last Assimilation, as both rules state that [s] can be followed by [s] but not [ʃ], and [ʃ] can be followed by [ʃ] but not [s]. The only difference is the environments of these restrictions, as shown in table 2.

From both a linguistic and a cognitive perspective, First-Last Assimilation seems plausible not only because long-distance dependencies between sounds are attested in natural language, but also because word edges have special status in phonology (Beckman 1998, Endress, Nespor, and Mehler 2009). Sounds at these positions are usually more perceptually salient, and some phonological rules are edge-sensitive. In this light, First-Last Assimilation is not such a strange pattern.

2.4 C'Lela

Another reason to think that First-Last Assimilation is not a strange pattern is that it is very similar to an attested pattern: a vowel harmony pattern in c'Lela (Dettweiler 2000, Pulleyblank 2002, Archangeli and Pulleyblank 2007). C'Lela is a Niger-Congo language, spoken in Nigeria. The direct object 1st person pronoun [-mi]/[-me] alternates depending on the height of the root vowel. If the vowel in the root is high, the suffix [-mi] surfaces, as in (1). If the vowel in the root is nonhigh, [-me] surfaces, as in (2). (Examples from Archangeli and Pulleyblank 2007:359.)

(1) [buzᵊkᵊ-mi] 'chased me'

(2) [ɛpkᵊ-me] 'bit me'

C'Lela allows suffix stacking, and interestingly, if there is more than one suffix attached to a root, only the final suffix assimilates to the vowel in the root. The word-medial suffix becomes transparent. Consider the following examples (from Archangeli and Pulleyblank 2007:360):

(3) *High root with single suffix*
a. i-z**i**s-**i** 'CM-long-CM'
b. u-p**u**s-**u** 'CM-white-CM'

(4) *High root with two suffixes*
a. i-z**i**s-i-n**i** 'CM-long-CM-ADJM'
b. u-p**u**s-u-n**i** 'CM-white-CM-ADJM'

(5) *Nonhigh root with single suffix*
a. i-r**e**k-**e** 'CM-small-CM'
b. ugʲ**ɔ**z-**o** 'CM-red-CM'

(6) *Nonhigh root with two suffixes*
a. i-r**e**k-i-n**e** 'CM-small-CM-ADJM'
b. u-gʲ**ɔ**z-u-n**e** 'CM-red-CM-ADJM'

The vowels in the high roots and their class marker suffixes in examples (3a–b) all agree in height. This is also the case for nonhigh roots, as in examples (5a–b). When an additional adjectival suffix is attached to the stems in (3a–b), the class marker suffixes remain high, as in (4a–b). However, when an additional adjectival suffix is attached to the stems in (5a–b), medial suffixes surface as [-i] and [-u], as in (6a–b). The newly added final suffixes still surface as nonhigh vowels, and therefore only the vowels in the root and the final suffix agree in height.

One interpretation of the above data is that they represent an edge-sensitive vowel harmony pattern. However, it should be noted that prefixes do not seem to participate in the vowel harmony process in c'Lela, as shown in examples (5)–(6). In addition, words with multisyllabic roots are limited, and therefore no example of a root with vowels of different height (with the exception of [-ᵊ], analyzed as a nonphonemic featureless mora (Pulleyblank 2002:260)) was found. From these examples, one can only conclude that the trigger of the vowel harmony in c'Lela is morphologically bound (the vowel is in a root) and the target is position-bound (the final suffix). This is different from First-Last Assimilation, in which both the trigger and the target are position-bound. However, if one assumes that the c'Lela harmony pattern is indeed a case of First-Last Assimilation, it is still true that First-Last Assimilation is much rarer than Sibilant Harmony patterns. One reviewer suggests that once a morpheme-bound trigger and a position-bound target are admitted as elements of a theory, then formally, any combination of those ''bindings'' is possible and hence the First-Last Assimilation system should also be possible. However, this conclusion is predicated upon the particular formal mechanisms available to the theory. This is exactly the issue being addressed here. The results reported in sections 4–5 suggest that either no such mechanism ought to exist or even if the combination is formally admissible, it is rarely utilized, as evidenced by c'Lela's outlier status. This is exactly what the Subregular Hypothesis predicts: this type of pattern is psychologically significantly more complex.

3 Evaluating the Subregular Hypothesis

3.1 Previous Literature on the Learnability of Patterns

A body of research has shown that both children and adults can learn phonotactic patterns by extracting the regularities exhibited in an artificial language (e.g., Infant studies: Chambers, Onishi, and Fisher 2003, Seidl et al. 2009; Adult studies: Dell et al. 2000, Onishi, Chambers, and Fisher 2002, Wilson 2003, Goldrick 2004, Peperkamp, Skoruppa, and Dupoux 2006, Finley and Badecker 2009a,b, Finley 2011, 2012, Koo and Callahan 2012). Each pattern examined in the previous studies falls into one of two categories: adjacent dependencies or nonadjacent dependencies. A comprehensive review on using the artificial-language-learning paradigm with children and adults (Folia et al. 2010) states that studies investigating such paradigms with both adults and children generally report similar or ''equivalent'' findings for both groups.

The well-formedness of adjacent dependencies can be judged by the cooccurrence of contiguous segments. In other words, these patterns are Strictly Local. The patterns studied by Aslin, Saffran, and Newport (1998), Dell et al. (2000), Onishi, Chambers, and Fisher (2002), Chambers, Onishi, and Fisher (2003), and Goldrick (2004) all fall into this category, and the length of the relevant substrings does not exceed 2. The results of these studies indicate that humans can learn Strictly Local languages in an artificial-language-learning setting.

Nonadjacent dependencies are also readily learned (Pycha et al. 2003, Wilson 2003, Newport and Aslin 2004, Onnis et al. 2005, Finley and Badecker 2009a,b, Finley 2011, 2012). The patterns tested in these studies require some segment x to agree with some segment y, and the two agreeing segments can be separated by a number of nonparticipating segments. These patterns are in fact Strictly Piecewise patterns. Thus, evidence from the research cited so far is consistent with the Subregular Hypothesis.

Koo and Callahan's (2012) study differs from the studies mentioned so far, as the long-distance dependency pattern in their study could be interpreted as position-bound. The pattern they studied can be understood as requiring the adult participants to learn the probability of cooccurrence of the first and last consonants of words with three consonants. All of the words in Koo and Callahan's experiments were trisyllabic, with the structure CVCVCV. The language presented to the participants can be described by the following two rules:

1. Whenever [s] is the onset of the first syllable, [l] cannot be the onset of the last syllable.
2. Whenever [l] is the onset of the first syllable, [m] cannot be the onset of the last syllable.

These two rules were consistent with the First-Last Assimilation pattern, except that they describe an arbitrary dependency pattern rather than assimilation.[2]

Under these rules, the sounds [s] and [l], and [l] and [m], cannot cooccur at a distance, but they can be adjacent to each other on the consonant tier. This pattern was shown to be learnable under Koo and Callahan's experimental settings. The significance of this finding for the Subregular Hypothesis is discussed below.

3.2 Evaluation by Comparison

It is impossible to provide empirical proof that a particular pattern is not learnable. For example, suppose the results of an artificial-language-learning experiment indicate that a pattern was not learned by its participants. This null result is insufficient to prove the pattern is unlearnable since there might be another paradigm (say, one with a longer training time) that might give different results. Therefore, the study reported here instead tested a weaker version of the Subregular Hypothesis: First-Last Assimilation is harder to learn than Sibilant Harmony.

There are no studies that directly compare the learnability of patterns that belong to these two classes and one that does not. Such a comparison can generate four logically possible outcomes: (1) both patterns are learned; (2) Sibilant Harmony is learned, while First-Last Assimilation is not; (3) neither pattern is learned; (4) First-Last Assimilation is learned, while Sibilant Harmony is not. The possible outcomes are summarized in table 3.

All of these scenarios except for outcome 4 are compatible with the Subregular Hypothesis. Therefore, just demonstrating that a non-Subregular pattern is learnable under some artificial conditions is not sufficient to reject this hypothesis. This is why Koo and Callahan's (2012) study, while interesting, does not falsify the Subregular Hypothesis. However, an experimental paradigm that produces outcome 4 would falsify the Subregular Hypothesis. Additionally, showing that both Subregular and non-Subregular patterns are learnable or unlearnable (outcomes 1 and 3) is

[2] As Koo and Callahan point out, these rules are also consistent with Tier-Based Strictly Local and Strictly Piecewise patterns with a window size of 3.

Table 3
Logically possible experimental outcomes that could be obtained from comparing the learnability of a Subregular pattern with that of a non-Subregular pattern

Paradigms	Subregular (Strictly Local/Strictly Piecewise/Tier-Based Strictly Local) e.g., Sibilant Harmony	Non-Subregular (Non–Strictly Local/Strictly Piecewise/Tier-Based Strictly Local) e.g., First-Last Assimilation
1	Learned	Learned
2	Learned	Not learned
3	Not learned	Not learned
4	Not learned	Learned

not particularly informative. The finding that both patterns are unlearnable in an experimental context would be unexpected; since Subregular patterns are found in natural languages, not being able to learn them would suggest a faulty experimental design. On the other hand, if both patterns are learnable, the Subregular Hypothesis is not rejected but also not supported because the result fails to support or reject the hypothesis that Subregular patterns are easier to learn than the non-Subregular pattern. This result also indicates a faulty experimental design, because it is too easy to learn both kinds of patterns. Outcome 2 can be interpreted as evidence in favor of the Subregular Hypothesis.

In order to compare the learnability of two patterns, the patterns must differ as minimally as possible, and the paradigm must give equal training to each experimental condition. As explained above, First-Last Assimilation and Sibilant Harmony are well-matched in many respects. The decision to test the learnability of First-Last Assimilation was not arbitrary. This pattern was chosen not only because it is a Regular but non-Subregular pattern, but also because it is very similar to Sibilant Harmony, an attested pattern. Computationally, the required memory is the same; only the pattern template is different (see table 2). Thus, the learnability of these two patterns can be compared fairly. Additionally, the first and last positions of a word are both privileged in terms of saliency and are relevant in phonology. Finally, the existence in c'Lela of a pattern that resembles First-Last Assimilation can plausibly be interpreted as evidence for its learnability. All these properties of First-Last Assimilation make it a good candidate for evaluating the Subregular Hypothesis.

4 Experiment 1

4.1 Method

The hypothesis of this study is that the absence of certain types of phonological patterns in the world's languages is due to limitations on what the phonological learner can extrapolate from the speech input. This hypothesis was tested empirically in two artificial-language-learning experiments.

4.1.1 Subjects Sixty-six monolingual adult native speakers of American English were recruited for experiment 1. Participants were students from the University of Delaware, between 18 and 27 years of age. They were compensated for their participation with either course credit or $10.

4.1.2 Procedure The experiment took place in a soundproof booth in the Phonetics and Phonology Laboratory at the University of Delaware. The experiment consisted of two experimental conditions (Sibilant Harmony and First-Last) and a control condition.

The procedure for both experimental conditions consisted of two phases: a training phase and a testing phase. The total duration for both training and testing was about 25 minutes. During the training phase of the two experimental conditions, participants listened to words that conformed to either a Sibilant Harmony or a First-Last Assimilation grammar (depending on the experimental condition) and were instructed to repeat each word orally after it was presented. The training contained 200 tokens (40 words $\times$ 5 repetitions) and the duration was approximately 15 minutes.

In the control condition, no training was given; participants were only given the test. In this condition, an alternative option would have been to provide participants with randomized training data (data that do not form any pattern). However, Finley and Badecker (2009a,b) found no differences between no-training control conditions and random training control conditions.

In the two experimental conditions, training was followed by a testing phase in which participants were presented with pairs of words and were asked to judge whether the first word or the second word of each pair was more likely to belong to the artificial language they had just heard during the training. Participants in the control condition were asked to judge whether they liked the first word or the second word of each pair better as a possible word. There were 48 pairs of novel test items, and the test took about 10 minutes to complete. All participants, regardless of condition, were given the same test with the same 48 pairs of test items.

4.1.3 Stimuli All training and test items were trisyllabic, with structure CV.CV.CVC. The only consonants used in constructing the items were [k, s, ʃ], and the only vowels were [a, ɛ, i, ɔ, u]. Half of the training items had a stop as the second consonant, and the other half had a stop as the third consonant. The first and last consonants were always sibilants.

In the Sibilant Harmony condition, all training items conformed to Sibilant Harmony. In the First-Last condition, all training items conformed to First-Last Assimilation.

Table 4 summarizes the types of training stimuli used. A complete list of stimuli is given in appendixes A and B.

One-third of the test items contained disagreeing sibilants as the first and last consonants (e.g., [s . . . s . . . ʃ]); these words conform to neither the Sibilant Harmony nor the First-Last Assimilation grammar (i.e., they are examples of *FL/*SH). Another third of the test items contained agreeing sibilants throughout (e.g., [s . . . s . . . s]); these words conform to both Sibilant Harmony and First-Last Assimilation (FL/SH). The final third contained agreeing sibilants only as the first and last consonants (e.g., [s . . . ʃ . . . s]); these words conform only to First-Last Assimilation (FL/*SH). The fourth logically possible type, words that conform to Sibilant Harmony but not First-Last Assimilation (*FL/SH), could not be instantiated, because all items that conform to Sibilant Harmony must also conform to First-Last Assimilation.

Table 4
Types of training items used in the Sibilant Harmony, First-Last, and control conditions. Vowels are omitted. (No training took place in the control condition.)

	Conditions	
Sibilant tier	Sibilant Harmony	First-Last
[s . . . s . . . s]	[s . . . k . . . s . . . s] [s . . . s . . . k . . . s]	[s . . . k . . . s . . . s] [s . . . s . . . k . . . s]
[ʃ . . . ʃ . . . ʃ]	[ʃ . . . k . . . ʃ . . . ʃ] [ʃ . . . ʃ . . . k . . . ʃ]	[ʃ . . . k . . . ʃ . . . ʃ] [ʃ . . . ʃ . . . k . . . ʃ]
[s . . . ʃ . . . s]	None	[s . . . k . . . ʃ . . . s] [s . . . ʃ . . . k . . . s]
[ʃ . . . s . . . ʃ]	None	[ʃ . . . k . . . s . . . ʃ] [ʃ . . . s . . . k . . . ʃ]

A two-alternative forced-choice design was used; the three types of test stimuli were pitted against each other and generated three types of pairings:

1. FL/*SH (FL only) vs. *FL/*SH (neither FL nor SH). For example, [s . . . ʃ . . . s] vs. [s . . . s . . . ʃ].[3]
2. FL/SH (FL and SH) vs. *FL/*SH (neither FL nor SH). For example, [s . . . s . . . s] vs. [s . . . s . . . ʃ].[4]
3. FL/*SH (FL only) vs. FL/SH (FL and SH). For example, [s . . . ʃ . . . s] vs. [s . . . s . . . s].[5]

4.1.4 Recording of Stimuli Natural stimuli were used for the experiments. A native speaker of Mandarin Chinese, a graduate student with phonetic training who was unaware of the experiments' purpose, was recruited to record the stimuli. Explicit training was given to the recorder to ensure that all stimuli were produced consistently. All vowels were pronounced as full vowels. Word stress (with the acoustic correlates of increased pitch and loudness) was placed on the penultimate syllable of all words, and the sibilant [ʃ] was pronounced with rounded lips.

4.1.5 Predictions The experiment was designed to investigate whether the choice made by participants in the test phase was influenced by the type of grammar they were exposed to in training. Table 5 summarizes the responses predicted if Sibilant Harmony and First-Last Assimilation were successfully learned in the respective conditions.

The results from both the Sibilant Harmony and First-Last groups were compared with those of the control group. Assuming that the control group should have no preference for either item

[3] The order of presentation of each word in a pair was counterbalanced, so this also includes *FL/*SH vs. FL/*SH.
[4] This also includes *FL/*SH vs. FL/SH.
[5] This also includes FL/SH vs. FL/*SH.

Table 5
Predicted preferences for each test pairing if Sibilant Harmony and First-Last Assimilation grammars were internalized

Conditions	Pairs: FL/*SH vs. *FL/*SH (e.g., [s . . . ʃ . . . s] vs. [s . . . s . . . ʃ])	FL/SH vs. *FL/*SH (e.g., [s . . . s . . . s] vs. [s . . . s . . . ʃ])	FL/SH vs. FL/*SH (e.g., [s . . . s . . . s] vs. [s . . . ʃ . . . s])
SH	No preference	[s . . . s . . . s] > [s . . . s . . . ʃ]	[s . . . s . . . s] > [s . . . ʃ . . . s]
FL	[s . . . ʃ . . . s] > [s . . . s . . . ʃ]	[s . . . s . . . s] > [s . . . s . . . ʃ]	No preference
Control	No preference	No preference	No preference

in each pairing (since no training was given), the results predicted for each experimental group if the participants successfully internalized the grammar they were exposed to are as shown in table 6.

4.2 Results

The descriptive statistics for the rates of choosing FL/*SH and FL/SH in all three types of test pairings are summarized in table 7.

Participants' responses were collected with the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and were modeled using a linear mixed-effects model with a binomial function. (The distribution of the test results being binomial because of the nature of a two-alternative forced-choice task, more traditional analyses using the *t*-test or ANOVA, which assume normally distributed data, are inappropriate.) The model was fitted in R (v.2.13.1) (R Development Core Team 2009), using the *lmer( )* function from the *lme4* package (Bates, Maechler, and Bolker 2011) for mixed-effects models. The model contained a fixed-effect Condition with three levels (control, Sibilant Harmony, and First-Last) and two random effects, Subject and Trial. For each

Table 6
Predicted results with respect to the control group for each test pairing if Sibilant Harmony and First-Last Assimilation grammars were internalized

Conditions	Pairs: FL/*SH vs. *FL/*SH (e.g., [s . . . ʃ . . . s] vs. [s . . . s . . . ʃ]) Rate of FL/*SH	FL/SH vs. *FL/*SH (e.g., [s . . . s . . . s] vs. [s . . . s . . . ʃ]) Rate of FL/SH	FL/SH vs. FL/*SH (e.g., [s . . . s . . . s] vs. [s . . . ʃ . . . s]) Rate of FL/SH
SH	~ Control	> Control	> Control
FL	> Control	> Control	~ Control

Table 7
Descriptive statistics for the control, Sibilant Harmony, and First-Last conditions

	Conditions		
	Control	Sibilant Harmony	First-Last
FL/*SH vs. *FL/*SH			
Mean rate of FL/*SH *(SE)*	0.49 (*0.03*)	0.49 (*0.03*)	0.52 (*0.03*)
FL/SH vs. *FL/*SH			
Mean rate of FL/SH *(SE)*	0.48 (*0.03*)	0.62 (*0.03*)	0.63 (*0.03*)
FL/SH vs. FL/*SH			
Mean rate of FL/*SH *(SE)*	0.45 (*0.03*)	0.56 (*0.03*)	0.58 (*0.03*)

analysis, the control condition was coded as the reference level, which was shown as the intercept in the output. With this set-up, the responses of the participants in each experimental condition could be compared directly with those of the participants in the control condition. Each model was compared with the empty model, where the fixed effect was replaced by 1. The function *anova( )* was used to perform a likelihood ratio test between the empty model and the respective individual model to check whether Condition was an important factor in its own right in each model.

The results for each type of pairing were analyzed separately because each pairing had a different dependent variable. The results were analyzed by examining the rate of choosing one type of stimuli over the other within a pairing. For example, in pairing 1, where FL/*SH was pitted against *FL/*SH, the rate of choosing FL/*SH was analyzed: a subject's response was coded as 1 if he or she chose the FL/*SH item and 0 otherwise. For pairings 2 and 3, a subject's response was coded as 1 if he or she chose the FL/SH item and 0 otherwise.

In the analysis, the 1-tailed test for cases in which the results were expected to be ‘‘Higher than control’’ was used. For cases in which ‘‘Same as control’’ was predicted, the 2-tailed test was used.

The mean rates of choosing FL/*SH when participants were presented with the FL/*SH vs. *FL/*SH pairings in all three conditions are shown in figure 2.

The likelihood ratio tests showed that only two out of three models with the fixed factor Condition were significantly different from their respective empty models. The first model, FL/*SH vs. *FL/*SH, was not significantly different from its empty model ($\chi^2 = 1.05$, $p = .59$), which means that Condition is not an important predictor in this model. The second model, FL/SH vs. *FL/*SH, and the third model, FL/SH vs. FL/*SH, were both significantly different from their empty models ($\chi^2 = 14.22$, $p < .001$ and $\chi^2 = 10.71$, $p = .005$, respectively). This means that Condition is an important factor in its own right in both models.

The model for the FL/*SH vs. *FL/*SH pairings showed that neither the Sibilant Harmony group's nor the First-Last group's responses differed significantly from the control group's (shown as *Intercept* in table 8). The log odds of the Sibilant Harmony participants' choosing FL/*SH

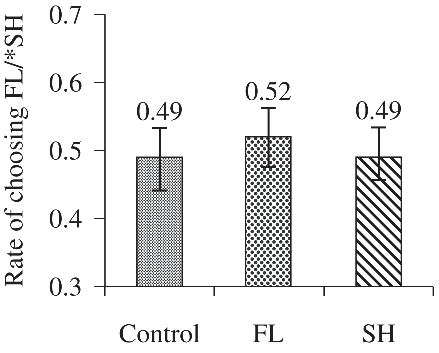

Figure 2
Mean rates of choosing FL/*SH when participants were presented with the pair FL/*SH vs. *FL/*SH ($N = 66$)

was not significantly higher than the log odds of the control participants' doing so ($p_{1\text{-}tailed}$ = .47), nor was the log odds of the First-Last participants' choosing FL/*SH significantly different from the log odds of the control participants' doing so ($p_{2\text{-}tailed}$ = .20).

The mean rates of choosing FL/SH when participants were presented with the FL/SH vs. *FL/*SH pairings in all three conditions are shown in figure 3.

As table 9 shows, the model for the FL/SH vs. *FL/*SH pairings suggests that the log odds of the Sibilant Harmony participants' choosing FL/SH was significantly higher than the log odds of the control participants' doing so ($p_{1\text{-}tailed} < .001$), and that the log odds of the First-Last participants' choosing FL/SH was also significantly higher than the log odds of the control participants' doing so ($p_{1\text{-}tailed} < .001$).

The mean rates of choosing FL/SH when participants were presented with the FL/SH vs. FL/*SH pairings in all three conditions are shown in figure 4.

As table 10 shows, the model for the FL/SH vs. FL/*SH pairings suggests that the log odds of the Sibilant Harmony participants' choosing FL/SH was significantly higher than the log odds of the control participants' doing so ($p_{1\text{-}tailed}$ = .004), and that the log odds of the First-Last

Table 8
Estimates of the conditions in the analysis of participants' responses in the pairing FL/*SH vs. *FL/*SH

*FL/*SH vs. *FL/*SH*	Estimate	Standard error	z	$p_{2\text{-}tailed}$	$p_{1\text{-}tailed}$
(Intercept)	−0.04638	0.14192	−0.327	.744	.372
Condition: SH	−0.0119	0.16435	−0.072	.942	.471
Condition: FL	0.14139	0.16439	0.860	.390	.195

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

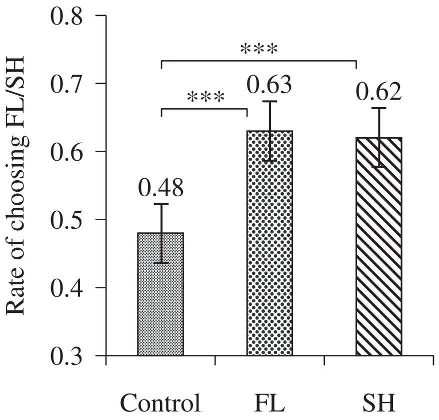

Figure 3
Mean rates of choosing FL/SH when participants were presented with the pair FL/SH vs. *FL/*SH ($N = 66$)

Table 9
Estimates of the conditions in the analysis of participants' responses in the pairing FL/SH vs. *FL/*SH

*FL/SH vs. *FL/*SH*	Estimate	Standard error	*z*	$p_{2\text{-tailed}}$	$p_{1\text{-tailed}}$
(Intercept)	−0.07203	0.16927	−0.426	.670436	.335218
Condition: SH	0.58131	0.18073	3.216	.001298**	.000649***
Condition: FL	0.6581	0.18134	3.629	.000284***	.000142***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

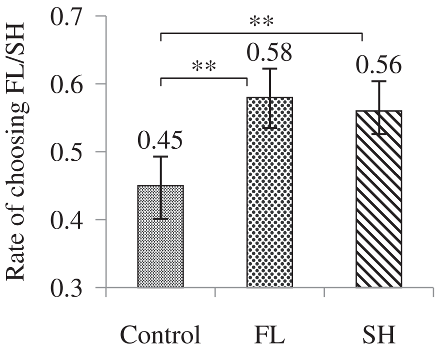

Figure 4
Mean rates of choosing FL/SH when participants were presented with the pair FL/SH vs. FL/*SH ($N = 66$)

Table 10
Estimates of the conditions in the analysis of participants' responses in the pairing FL/SH vs. FL/*SH

*FL/SH vs. FL/*SH*	Estimate	Standard error	z	$p_{2\text{-}tailed}$	$p_{1\text{-}tailed}$
(Intercept)	−0.2037	0.1526	−1.335	.18196	.09098
Condition: SH	0.4544	0.1697	2.678	.0074**	.0037**
Condition: FL	0.5394	0.1700	3.172	.00151**	.000755***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

participants' choosing FL/SH was significantly higher than the log odds of the control participants' doing so ($p_{2\text{-}tailed} = .002$).

A separate analysis was run to test whether the First-Last group performed significantly differently from the Sibilant Harmony group. The Sibilant Harmony group was coded as the reference group (*Intercept*) in this analysis, and since no specific direction was predicted for the results, 2-tailed tests were used. As table 11 shows, the responses of the First-Last group in all three types of test pairings do not differ significantly from those of the Sibilant Harmony group: FL/*SH vs. *FL/*SH ($p = .352$), FL/SH vs. *FL/*SH ($p = .676$), and FL/SH vs. FL/*SH ($p = .619$).

The results obtained match the predictions made by the Sibilant Harmony but not the First-Last Assimilation grammar. Therefore, it can be concluded that the Sibilant Harmony participants were able to internalize the Sibilant Harmony grammar.

Table 11
Estimates of the conditions in three types of test pairings with the Sibilant Harmony group as the reference group

*FL/*SH vs. *FL/*SH*	Estimate	Standard error	z	$p_{2\text{-}tailed}$
(Intercept)	−0.05818	0.14195	−0.41	.682
Condition: Control	0.01182	0.16435	0.072	.943
Condition: FL	0.15308	0.16441	0.931	.352
*FL/SH vs. *FL/*SH*	Estimate	Standard error	z	$p_{2\text{-}tailed}$
(Intercept)	0.50927	0.17142	2.971	.00297**
Condition: Control	−0.58127	0.18073	−3.216	.0013**
Condition: FL	0.07667	0.18317	0.419	.67552
*FL/SH vs. FL/*SH*	Estimate	Standard error	z	$p_{2\text{-}tailed}$
(Intercept)	0.25095	0.15285	1.642	.10063
Condition: Control	−0.45467	0.16968	−2.68	.00737**
Condition: FL	0.08469	0.17022	0.498	.61881

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The results for the First-Last condition were unexpected. Not only was the First-Last Assimilation grammar not learned by the First-Last participants, but their performance was not significantly different from that of the Sibilant Harmony participants in all three test pairings. When they were given the pairing *FL/*SH vs. FL/*SH, they did not perform significantly differently from the control group. If they had successfully learned the First-Last Assimilation grammar, their rate of choosing FL/*SH should have been higher than the control group's rate. When they were given the pairing *FL/*SH vs. FL/SH, their rate of choosing FL/SH was significantly higher than the control group's rate, a choice consistent with both the First-Last Assimilation and the Sibilant Harmony grammars. Finally, when they were given the pairing FL/*SH vs. FL/SH, their rate of choosing FL/SH was also significantly higher than the control group's rate; here, both items conform to First-Last Assimilation, but the participants showed a preference for the item that also conforms to Sibilant Harmony. In sum, the First-Last participants chose the items that conform to Sibilant Harmony significantly more than the items that do not, but they failed to choose items that conform only to First-Last Assimilation. Combining the results from all three pairings, we can conclude that the First-Last participants were unable to internalize the First-Last Assimilation grammar.

In addition, the First-Last participants were not expected to internalize the Sibilant Harmony grammar, because the First-Last condition training included items that do not conform to Sibilant Harmony (e.g., [s . . . ʃ . . . s] and [ʃ . . . s . . . ʃ]). Yet the First-Last participants seemed to ignore these words and chose to accept the Sibilant Harmony grammar anyway. It could be the case that the participants were heavily biased toward learning Sibilant Harmony and that the presence of stimuli that conform to both Sibilant Harmony and First-Last Assimilation led them to falsely assume the Sibilant Harmony grammar. Thus, as a follow-up, an additional experiment was conducted to alleviate this potential Sibilant Harmony bias by replacing the ambiguous FL/SH words with words that conform only to First-Last Assimilation and not to Sibilant Harmony (i.e., FL/*SH).

4.3 Discussion

The experiment in this study was designed to test the learnability of two phonotactic patterns in the fairest possible way. The learnability of two patterns, differing only minimally in phonological terms but differing in their computational characterizations, was compared. The results have shown that Sibilant Harmony was readily learned by humans in this paradigm. A mere 15 minutes of exposure to the grammar was sufficient to significantly affect the participants' behavior. The performance of the Sibilant Harmony participants matched the predictions in all three types of test-pairings and therefore provides strong evidence that the Sibilant Harmony grammar was internalized.

These results were expected, as Sibilant Harmony is both attested and belongs to the Strictly Piecewise pattern. On the other hand, participants who were exposed to the First-Last Assimilation grammar did not perform according to the predictions. The only way to establish whether the First-Last Assimilation grammar was learned is to examine the participants' overall performance in all three pairings. Since only the results for one pairing concur with the First-Last Assimilation grammar's prediction, there was insufficient evidence to claim that the First-Last Assimilation grammar was successfully learned in this experiment.

It could be true that the First-Last Assimilation grammar would be learnable if the amount of training was increased or if the stimuli were presented in a different format or method. The crucial argument drawn from these results is that given the same experimental setting and the same amount of training, the Sibilant Harmony grammar was learned but the First-Last Assimilation grammar was not (see table 3). First-Last Assimilation was at least more challenging for the participants to internalize than Sibilant Harmony was.

Furthermore, the participants who were exposed to First-Last Assimilation performed very similarly to the participants who were exposed to Sibilant Harmony. It must be noted that the training set contained words that conform to both Sibilant Harmony and First-Last Assimilation (i.e., FL/SH; e.g., [s . . . s . . . s]); the proportion of such words was 50% of the entire set of training items. The remaining 50% consisted of words that did not conform to Sibilant Harmony. That means that for every word that could be construed as evidence for Sibilant Harmony, there was another word that was not consistent with Sibilant Harmony. First-Last participants' performance could be explained by a heavy Sibilant Harmony bias, which is influential enough to suppress the counterevidence. Pearl (2008) suggests that when children are faced with ambiguous linguistic input, they implement a filter that causes them to ignore information in the ambiguous data. If this theory can be extended to adults learning a new language in an experimental paradigm, the implementation of a filter would be a plausible explanation for why the First-Last participants ignored part of the training data.

Because of the apparent Sibilant Harmony bias exhibited by the participants in the First-Last condition, an experiment testing an additional condition, *Intensive First-Last*, was conducted.

5 Experiment 2

5.1 Method

In the Intensive First-Last condition, all the training items in the original First-Last condition that conform to both Sibilant Harmony and First-Last Assimilation were replaced with words that conform only to First-Last Assimilation. The purpose of testing this condition was to verify whether First-Last Assimilation could be learned if the Sibilant Harmony bias was alleviated by removing potentially distracting or ambiguous stimuli.

5.1.1 Subjects Another 22 monolingual speakers of American English were recruited for this condition. They were compensated the same way as participants in Experiment 1.

5.1.2 Procedures There was only one condition in this experiment: Intensive First-Last. The procedures were the same as those for Experiment 1. Participants were given audio training before they entered the test phase.

5.1.3 Stimuli The Intensive First-Last training stimuli were constructed similarly to the First-Last stimuli in terms of length, syllable structure, and the phoneme inventory used. Words that conform to both Sibilant Harmony and First-Last Assimilation (e.g., [s . . . s . . . s] and [ʃ . . . ʃ . . . ʃ]) were replaced by words that conform only to First-Last Assimilation (e.g., [s . . . ʃ . . . s] and [ʃ . . . s . . . ʃ]). Instead of the four types of training stimuli used in Experiment 1, only two

Table 12
Descriptive statistics for the Intensive First-Last condition

	FL/*SH vs. *FL/*SH (e.g., [s . . . ʃ . . . s] vs. [s . . . s . . . ʃ]) Rate of FL/*SH	FL/SH vs. *FL/*SH (e.g., [s . . . s . . . s] vs. [s . . . s . . . ʃ]) Rate of FL/SH	FL/SH vs. FL/*SH (e.g., [s . . . s . . . s] vs. [s . . . ʃ . . . s]) Rate of FL/SH
Mean	0.553977	0.414773	0.357955
Standard error	0.026532	0.026297	0.025588

types were used. The test used in the test phase was the same test used in the First-Last and Sibilant Harmony conditions in Experiment 1.

5.1.4 Predictions The results from this experiment were compared with the results from the control group in Experiment 1. Assuming that the control group should have no preference for either item in each pairing (since no training was given), the results of the Intensive First-Last condition are predicted to differ significantly from the results of the First-Last condition once the Sibilant Harmony learning bias has been removed.

5.2 Results

The results of the Intensive First-Last condition are summarized in table 12. They differ significantly from the results of the First-Last condition; see table 13.

Table 13
Estimates of the control, Sibilant Harmony, First-Last, and Intensive First-Last conditions

*FL/*SH vs. *FL/*SH*	Estimate	Standard error	z	$p_{2\text{-tailed}}$	$p_{1\text{-tailed}}$
(Intercept)	−0.04709	0.13903	−0.339	.7348	.3674
Condition: SH	−0.01173	0.15632	−0.075	.9402	.4701
Condition: FL	0.14076	0.15636	0.9	.368	.184
Condition: Intensive FL	0.27044	0.15673	1.726	.0844	.0422*
*FL/SH vs. *FL/*SH*	Estimate	Standard error	z	$p_{2\text{-tailed}}$	$p_{1\text{-tailed}}$
(Intercept)	−0.07222	0.16698	−0.433	.665367	.332684
Condition: SH	0.57972	0.18039	3.214	.00131**	.000655***
Condition: FL	0.6564	0.181	3.627	.000287***	.000144***
Condition: Intensive FL	−0.29479	0.17941	−1.643	.100352	.050176
*FL/SH vs. FL/*SH*	Estimate	Standard error	z	$p_{2\text{-tailed}}$	$p_{1\text{-tailed}}$
(Intercept)	−0.2045	0.1563	−1.309	.19063	.095315
Condition: SH	0.457	0.1763	2.592	.00955**	.004775**
Condition: FL	0.5418	0.1767	3.067	.00216**	.00108**
Condition: Intensive FL	−0.4116	0.1787	−2.303	.02128*	.01064*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

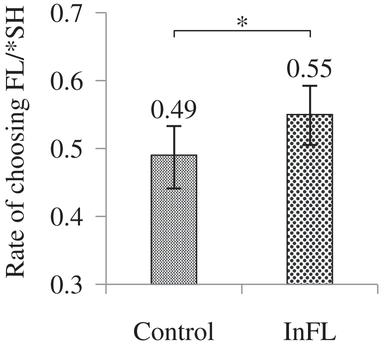

Figure 5
Mean rates of choosing FL/*SH when participants were presented with the pair FL/*SH vs. *FL/*SH in the Intensive First-Last condition ($N = 44$)

The rates of choosing FL/*SH when participants were presented with the pair FL/*SH vs. *FL/*SH in the Intensive First-Last condition are shown in figure 5; the rates of choosing FL/SH when participants were presented with the pair FL/SH vs. *FL/*SH are shown in figure 6; and the rates of choosing FL/SH when participants were presented with the pair FL/SH vs. FL/*SH are shown in figure 7.

The data obtained from the Intensive First-Last condition were added to the original data, and the whole dataset was regressed again with the same model. The fixed effect then consisted of four levels: control, Sibilant Harmony, First-Last, and Intensive First-Last. The estimates, z values, and p values of the three models (one for each type of test pairing) are included in table 13.

When the Intensive First-Last participants were given the pairing *FL/*SH vs. FL/*SH, the log odds of their choosing FL/*SH was significantly higher than the log odds of the control participants' doing so ($p_{1\text{-}tailed} = .04$). When the Intensive First-Last participants were given the pairing *FL/*SH vs. FL/SH, their log odds of choosing FL/SH went opposite to the prediction

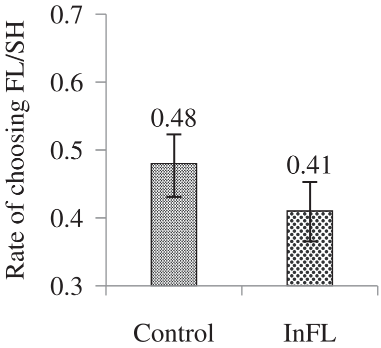

Figure 6
Mean rates of choosing FL/SH when participants were presented with the pair FL/SH vs. *FL/*SH in the Intensive First-Last condition ($N = 44$)

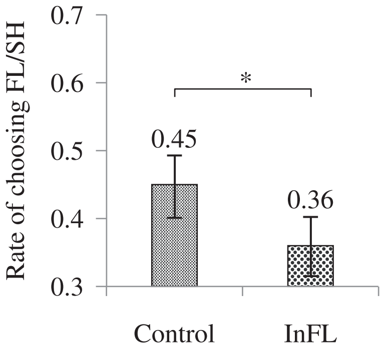

Figure 7
Mean rates of choosing FL/SH when participants were presented with the pair FL/SH vs. FL/*SH in the Intensive First-Last condition ($N = 44$)

made by the First-Last Assimilation grammar. For this reason, a post-hoc 2-tailed test was conducted to interpret the results. The analysis indicated that the log odds of the Intensive First-Last participants' choosing FL/SH does not differ significantly from the log odds of the control participants' doing so ($p_{2\text{-}tailed} = .100$). Finally, when the Intensive First-Last participants were given the pairing FL/SH vs. FL/*SH, their log odds of choosing FL/SH again went opposite to the prediction of the First-Last Assimilation grammar. A post-hoc 2-tailed test was conducted, and it confirmed that the Intensive First-Last participants' rate of choosing FL/SH, given this pairing, was significantly lower than the control participants' rate. Combining the results from all three pairings, we can conclude that participants in the Intensive First-Last condition only preferred stimuli that conform to First-Last Assimilation; they did not prefer stimuli that conform to Sibilant Harmony (i.e., FL/*SH).

5.3 Discussion

These results indicate that participants who were exposed only to FL/*SH stimuli during training internalized a rule different from First-Last Assimilation. Since all FL/*SH stimuli had the form [s . . . ʃ . . . s] or [ʃ . . . s . . . ʃ] (on the sibilant level), it is likely that these participants internalized a sibilant disharmony rule that requires neighboring sibilants to be disharmonic with respect to each other, a pattern that can be captured by a Tier-Based Strictly Local grammar. This could explain why participants did not prefer FL/SH (e.g., [s . . . s . . . s]) words. Nonetheless, Intensive First-Last participants definitely failed to internalize the First-Last Assimilation grammar that was intended in this study. Together, the Sibilant Harmony, First-Last, and Intensive First-Last condition results, all obtained in a carefully controlled experimental setting, show that First-Last Assimilation is harder to learn than Sibilant Harmony.

A possible follow-up could be designed to avoid the problem mentioned in the previous paragraph, of participants internalizing a sibilant disharmony rule. Instead of replacing all the training items that conform to both First-Last Assimilation and Sibilant Harmony (i.e., FL/SH words) with FL/*SH words, the proportion of FL/SH words in the training set could be lowered.

This would differentiate the First-Last Assimilation grammar from the disharmony grammar (as FL/SH is inconsistent with the disharmony rule), but the Sibilant Harmony bias would still be alleviated because of the fewer occurrences of FL/SH words.

Nonetheless, even if First-Last Assimilation was learned in this paradigm, the results would be meaningless unless Sibilant Harmony was *not* learned in the same paradigm (see table 3). In other words, the experiments conducted to date match the predictions of the Subregular Hypothesis.

6 Implications for Phonological Theories

This section discusses what the results of Experiments 1 and 2 mean for phonological theories.

For OT, the results provide another way to evaluate proposed constraints in CON. In fact, currently proposed markedness constraints can be studied with respect to the Subregular properties investigated here. For example, each of these constraints can be examined to see whether it falls into the Strictly Local or Strictly Piecewise class.

Local constraint conjunction (Smolensky 1995, 1997, 2005) allows simple constraints such as *#s, *ʃ#, *#ʃ, and *s# to be conjoined as *#s & *ʃ# and *#ʃ & *s#. These two sets of conjoined constraints will rule out any candidates that violate First-Last Assimilation. Therefore, to account for the absence of First-Last Assimilation in typology, the power of constraint conjunction must somehow be restricted within the theory.

Even if such conjoined constraints are omitted from CON, the question remains whether the First-Last Assimilation pattern can be derived from the interaction of simple constraints in OT. Answering this question is beyond the scope of this article, but it should be noted that the interaction of simple constraints can yield non-Regular patterns (Frank and Satta 1998, Riggle 2004, Gerdemann and Hulden 2012). Thus, it is known that more complex patterns can be obtained within OT through the interaction of simple constraints.

If the absence of First-Last Assimilation in typology is not accidental, then its absence poses a challenge for channel-biased explanations of typology (Ohala 1993, Blevins 2004). First-Last Assimilation can theoretically be derived from Sibilant Harmony as follows. The sibilants in the middle of the word are misheard, and therefore harmony fails to be maintained, but the sibilants at word edges are not misheard because they are more salient to listeners. As a result, Sibilant Harmony is maintained only at these word edge positions, which is the First-Last Assimilation pattern. In other words, the Sibilant Harmony pattern is a precursor to First-Last Assimilation. If the Subregular Hypothesis is correct, as the research reported here suggests, how best to incorporate it into phonological theory remains an interesting question for future research.

7 Conclusions

The experimental results of this study have provided empirical evidence for the difference in learnability of two carefully matched phonotactic patterns. The Sibilant Harmony pattern is an attested long-distance dependency pattern that belongs to the Strictly Piecewise class. First-Last Assimilation, on the other hand, is an unattested, non–Strictly Local, non–Strictly Piecewise, non–Tier-Based Strictly Local but Regular pattern. The learnability of these two patterns was compared, and the results suggest that Sibilant Harmony was learnable in the experimental paradigm used in this study, while First-Last Assimilation was not. The results concur with the hypothesis that phonotactic patterns that reside outside of the Strictly Local, Strictly Piecewise, and Tier-Based Strictly Local classes are less easily learned than those that reside within them. These findings imply that the computational boundaries proposed by the Subregular Hypothesis are psychologically real.

Appendix A: Training Stimuli

The training stimuli used in the Sibilant Harmony (SH), First-Last (FL), and Intensive First-Last (Intensive FL) conditions are included here.

Conditions	Items			
SH	ʃaʃɛkuʃ	sasɔkɔs	ʃakuʃuʃ	sakɛsas
	ʃɛʃukuʃ	sɛsukɔs	ʃɛkɛʃuʃ	sɛkɔsɛs
	ʃiʃɔkɛʃ	sisɛkɔs	ʃikiʃɔʃ	sikisis
	ʃɔʃɔkɛʃ	sɔsakɔs	ʃɔkaʃɛʃ	sɔkusas
	ʃuʃakiʃ	susukɔs	ʃukɔʃɔʃ	sukasus
	ʃaʃɔkuʃ	sasakus	ʃakuʃiʃ	sakisɔs
	ʃɛʃɔkiʃ	sɛsukis	ʃɛkɛʃɔʃ	sɛkɛsɛs
	ʃiʃukɔʃ	sisɔkɔs	ʃikɛʃaʃ	sikasɔs
	ʃɔʃɔkuʃ	sɔsikas	ʃɔkɛʃiʃ	sɔkɔsis
	ʃuʃakɛʃ	susɔkis	ʃukɔʃɛʃ	sukɛsas
FL	ʃasɔkuʃ	saʃakus	ʃaʃɛkuʃ	sasɔkɔs
	ʃɛsɔkiʃ	sɛʃukis	ʃɛʃukuʃ	sɛsukɔs
	ʃisukɔʃ	siʃɔkɔs	ʃiʃɔkɛʃ	sisɛkɔs
	ʃɔsɔkuʃ	sɔʃikas	ʃɔʃɔkɛʃ	sɔsakɔs
	ʃusakɛʃ	suʃɔkis	ʃuʃakiʃ	susukɔs
	ʃakusiʃ	sakiʃɔs	ʃakuʃuʃ	sakɛsas
	ʃɛkɛsɔʃ	sɛkɛʃɛs	ʃɛkɛʃuʃ	sɛkɔsɛs
	ʃikɛsaʃ	sikaʃɔs	ʃikiʃɔʃ	sikisis
	ʃɔkɛsiʃ	sɔkɔʃis	ʃɔkaʃɛʃ	sɔkusas
	ʃukɔsɛʃ	sukɛʃas	ʃukɔʃɔʃ	sukasus
Intensive FL	ʃasɔkuʃ	saʃakus	ʃakusiʃ	sakiʃɔs
	ʃɛsɔkiʃ	sɛʃukis	ʃɛkɛsɔʃ	sɛkɛʃɛs
	ʃisukɔʃ	siʃɔkɔs	ʃikɛsaʃ	sikaʃɔs
	ʃɔsɔkuʃ	sɔʃikas	ʃɔkɛsiʃ	sɔkɔʃis
	ʃusakɛʃ	suʃɔkis	ʃukɔsɛʃ	sukɛʃas
	ʃasɛkuʃ	saʃɔkɔs	ʃakusuʃ	sakɛʃas
	ʃɛsukuʃ	sɛʃukɔs	ʃɛkɛsuʃ	sɛkɔʃɛs
	ʃisɔkɛʃ	siʃɛkɔs	ʃikisɔʃ	sikiʃis
	ʃɔsɔkɛʃ	sɔʃakɔs	ʃɔkasɛʃ	sɔkuʃas
	ʃusakiʃ	suʃukɔs	ʃukɔsɔʃ	sukaʃus

Appendix B: Test Stimuli

The test stimuli used in all conditions are included here.

FL/*SH vs. *FL/*SH		FL/SH vs. *FL/*SH		FL/SH vs. FL/*SH	
Word 1	Word 2	Word 1	Word 2	Word 1	Word 2
sɛkɔʃɔs	ʃɛkɔʃɔs	sukisas	sukisaʃ	ʃɔʃukɔʃ	ʃɔsukɔʃ
ʃasɔkaʃ	ʃasɔkas	ʃɛʃɛkaʃ	sɛʃɛkaʃ	sukɛsus	sukɛʃus
suʃɛkɔs	suʃɛkɔʃ	ʃɔkuʃiʃ	ʃɔkuʃis	sisakus	siʃakus
ʃikisaʃ	sikisaʃ	sisɔkus	ʃisɔkus	ʃakaʃɔʃ	ʃakasɔʃ
sɛʃɔkɔs	sɛʃɔkɔʃ	ʃɛkɛʃaʃ	ʃɛkɛʃas	ʃɔkuʃɔʃ	ʃɔkusɔʃ
ʃakɔsaʃ	sakɔsaʃ	ʃɔʃukiʃ	sɔʃukiʃ	susɛkus	suʃɛkus
sukɛʃɔs	ʃukɛʃɔs	sikɔsus	sikɔsuʃ	sikasus	sikaʃus
ʃisikaʃ	ʃisikas	susikas	ʃusikas	ʃaʃakɔʃ	ʃasakɔʃ
ʃikuʃis	sikuʃis	sɔkasiʃ	sɔkasis	sɔʃakas	sɔsakas
ʃɛsakis	ʃɛsakiʃ	ʃikɔʃis	ʃikɔʃiʃ	ʃusɛkiʃ	ʃuʃɛkiʃ
sɔkisɔʃ	ʃɔkisɔʃ	ʃɛsikɔs	sɛsikɔs	sakuʃɛs	sakusɛs
saʃɛkɛʃ	saʃɛkɛs	saʃikuʃ	ʃaʃikuʃ	ʃɛkɔsaʃ	ʃɛkɔʃaʃ
ʃɔsikɔs	ʃɔsikɔʃ	ʃɔsakis	sɔsakis	saʃukɛs	sasukɛs
siʃukiʃ	siʃukis	siʃɔkiʃ	ʃiʃɔkiʃ	ʃɛsɔkaʃ	ʃɛʃɔkaʃ
sɛkasiʃ	ʃɛkasiʃ	sɛkisɔʃ	sɛkisɔs	sɔkaʃas	sɔkasas
ʃakɛʃɛs	sakɛʃɛs	ʃakiʃus	ʃakiʃuʃ	ʃukɛsiʃ	ʃukɛʃiʃ

Appendix C: Glossary of Abbreviations

This glossary provides definitions of the abbreviations used in the article.

FL *First-Last Assimilation* is a hypothetical long-distance dependency pattern that requires the initial and final segments of a word to be harmonic. It is a Regular but non–Strictly Local, non–Strictly Piecewise, and non–Tier-Based Strictly Local pattern.

SH *Sibilant Harmony* is an attested long-distance dependency pattern that requires all of the sibilants within a word to agree in anteriority. It is a Strictly Piecewise pattern.

SL *Strictly Local* languages make up a proper subset of the Regular languages. Strictly Local patterns are those that can be described in terms of a finite set of forbidden (contiguous) sequences of symbols of length k (thus, this pattern is called Strictly k-Local). Informally, the term *Strictly Local* refers to local dependency patterns.

SP *Strictly Piecewise* languages make up a proper subset of the Regular languages. Strictly Piecewise languages make distinctions on the basis of (potentially discontiguous) subsequences of length k. A string is a subsequence of another string if and only if its symbols occur in the other string in order. Informally, the term *Strictly Piecewise* refers to long-distance dependencies.

TSL *Tier-Based Strictly Local* languages make up a proper subset of the Regular languages. Tier-Based Strictly Local patterns are essentially local dependency patterns operating over abstract phonological tiers.

References

Archangeli, Diana, and Douglas Pulleyblank. 2007. Harmony. In *The Cambridge handbook of phonology*, ed. by Paul de Lacy, 353–378. Cambridge: Cambridge University Press.

Aslin, Richard N., Jenny Saffran, and Elissa L. Newport. 1998. Computation of conditional probability statistics by human infants. *Psychological Science* 9:321–324.

Barnes, Jonathan. 2002. Positional neutralization: A phonologization approach to typological patterns. Doctoral dissertation, University of California, Berkeley.

Bates, Douglas, Martin Maechler, and Ben Bolker. 2011. lme4: Linear mixed-effects models using s4 classes [Computer software manual]. Available at http://CRAN.R-project.org/package = lme4.

Beckman, Jill. 1998. Positional faithfulness. Doctoral dissertation, University of Massachusetts, Amherst.

Berent, Iris, Colin Wilson, Gary F. Marcus, and Douglas K. Bemis. 2012. On the role of variables in phonology: Remarks on Hayes and Wilson 2008. *Linguistic Inquiry* 43:97–119.

Blevins, Juliette. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.

Chambers, Kyle E., Kristine H. Onishi, and Cynthia Fisher. 2003. Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87:B69–B77.

Dell, Gary S., Kristopher D. Reed, David R. Adams, and Antje S. Meyer. 2000. Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:1355–1367.

Dettweiler, Stephen H. 2000. Vowel harmony and neutral vowels in c'Lela. *Journal of West African Languages* 28:3–18.

Endress, Ansgar D., Marina Nespor, and Jacques Mehler. 2009. Perceptual and memory constraints on language acquisition. *Trends in Cognitive Science* 13:348–353.

Finley, Sara. 2011. The privileged status of locality in consonant harmony. *Journal of Memory and Language* 65:74–83.

Finley, Sara. 2012. Testing the limits of long-distance learning: Learning beyond the three-segment window. *Cognitive Science* 36:740–756.

Finley, Sara, and William Badecker. 2009a. Artificial grammar learning and feature-based generalization. *Journal of Memory and Language* 61:423–437.

Finley, Sara, and William Badecker. 2009b. Right-to-left biases for vowel harmony: Evidence from artificial grammar. In *NELS 38*, ed. by Anisa Schardl, Martin Walkow, and Muhammad Abdurrahman, 269–282. Amherst: University of Massachusetts, Graduate Linguistic Student Association.

Folia, Vasiliki, Julia Uddén, Meiou de Vries, Christian Forkstam, and Karl Magnus Petersson. 2010. Artificial language learning in adults and children. *Language Learning* 60:188–220.

Frank, Robert, and Giorgio Satta. 1998. Optimality Theory and the generative complexity of constraint violability. *Computational Linguistics* 24:307–315.

Gerdemann, Dale, and Mans Hulden. 2012. Practical finite state Optimality Theory. In *Proceedings of the 10th International Workshop on Finite State Methods and Natural Language Processing*, 10–19. Association for Computational Linguistics. Available at http://aclweb.org/anthology/W12-6202.

Goldrick, Matt. 2004. Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 51:586–603.

Goldsmith, John, and Jason Riggle. 2012. Information theoretic approaches to phonological structure: The case of Finnish vowel harmony. *Natural Language and Linguistic Theory* 20:859–896.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory. April*

26–27, 2003 at Stockholm Univ. Sweden, ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl, 113–122. Stockholm: Stockholm University.

Hale, Mark, and Charles Reiss. 2000. Substance abuse and ''dysfunctionalism'': Current trends in phonology. *Linguistic Inquiry* 31:157–169.

Hansson, Gunnar. 2001. Theoretical and typological issues in consonant harmony. Doctoral dissertation, University of California, Berkeley.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Heinz, Jeffrey. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41:623–661.

Heinz, Jeffrey. 2011a. Computational phonology part I: Foundations. *Language and Linguistics Compass* 5:140–152.

Heinz, Jeffrey. 2011b. Computational phonology part II: Grammars, learning, and the future. *Language and Linguistics Compass* 5:153–168.

Heinz, Jeffrey, Chetan Rawal, and Herbert Tanner. 2011. Tier-based strictly local constraints for phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon, USA*, 58–64. Available at http://www.aclweb.org/anthology/P/P11/P11-2011.pdf.

Kaplan, Ronald, and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20:331–378.

Koo, Hahn, and Lydia Callahan. 2012. Tier-adjacency is not a necessary condition for learning phonotactic dependencies. *Language and Cognitive Processes* 27:1425–1432.

McNaughton, Robert, and Seymour Papert. 1971. *Counter-free automata.* Cambridge, MA: MIT Press.

Moreton, Elliott. 2008. Learning bias as a factor in phonological typology. In *WCCFL 26: Proceedings of the 26th West Coast Conference on Formal Linguistics*, ed. by Charles B. Chang and Hannah J. Haynie, 393–401. Somerville, MA: Cascadilla Press.

Newport, Elissa L., and Richard N. Aslin. 2004. Learning at a distance I. Statistical learning of nonadjacent dependencies. *Cognitive Psychology* 48:127–162.

Nowak, Martin, Natalia Komarova, and Partha Niyogi. 2002. Computational and evolutionary aspects of language. *Nature* 417:611–617.

Ohala, John. 1993. The phonetics of sound change. In *Historical linguistics: Problems and perspectives*, ed. by Charles Jones, 237–278. Harlow: Longman.

Onishi, Kristine H., Kyle E. Chambers, and Cynthia Fisher. 2002. Learning phonotactic constraints from brief auditory experience. *Cognition* 83:B13–B23.

Onnis, Luca, Padraic Monaghan, Korin Richmond, and Nick Chater. 2005. Phonology impacts segmentation in online speech processing. *Journal of Memory and Language* 53:225–237.

Pearl, Lisa. 2008. Putting the emphasis on unambiguous: The feasibility of data filtering for learning English metrical phonology. In *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Child Language Development*, ed. by Harvey Chan, Heather Jacob, and Enkeleida Kapia, 390–401. Somerville, MA: Cascadilla Press.

Peperkamp, Sharon, Katrin Skoruppa, and Emmanuel Dupoux. 2006. The role of phonetic naturalness in phonological rule acquisition. In *BUCLD 30: Proceedings of the 30th annual Boston University Conference on Language Development*, ed. by David Bamman, Tatiana Magnitskaia, and Colleen Zaller, 464–475. Somerville, MA: Cascadilla Press.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar.* Cambridge, MA: Blackwell.

Pulleyblank, Douglas. 2002. Harmony drivers: No disagreement allowed. In *Proceedings of the Twenty-eighth Annual Meeting of the Berkeley Linguistics Society*, ed. by Julie Larson and Mary Paster, 249–267. Berkeley: University of California, Berkeley Linguistics Society.

Pycha, Anne, Pawel Nowak, Eurie Shin, and Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In *WCCFL 22: Proceedings of the 22nd West Coast Conference* *on Formal Linguistics*, ed. by Gina Garding and Mimu Tsujimura, 423–435. Somerville, MA: Cascadilla Press.

R Development Core Team. 2009. *R: A language and environment for statistical computing* [Computer software manual]. Available at http://www.R-project.org.

Riggle, Jason. 2004. Generation, recognition, and learning in finite state Optimality Theory. Doctoral dissertation, UCLA, Los Angeles, CA.

Rogers, James, Jeffrey Heinz, Gil Bailey, Matt Edlefsen, Molly Visscher, David Wellcome, and Sean Wibel. 2010. On languages piecewise testable in the strict sense. In *The mathematics of language*, ed. by Christian Ebert, Gerhard Jäger, and Jens Michaelis, 255–265. New York: Springer.

Rogers, James, Jeffrey Heinz, Margaret Fero, Jeremy Hurst, Dakotah Lambert, and Sean Wibel. 2013. Cognitive and sub-Regular complexity. In *Proceedings of the 17th Conference on Formal Grammar*, ed. by Glyn Morrill and Mark-Jan Nederhof, 90–108. Berlin: Springer.

Rogers, James, and Geoffrey K. Pullum. 2011. Aural pattern recognition experiments and the Subregular hierarchy. *Journal of Logic, Language and Information* 20:329–342.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80:475–531.

Sapir, Edward, and Harry Hoijer. 1967. *The phonology and morphology of the Navaho language.* Berkeley: University of California Press.

Seidl, Amanda, Alejandrina Cristià, Amelie Bernard, and Kristine H. Onishi. 2009. Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development* 5:191–202.

Smolensky, Paul. 1995. On the internal structure of the constraint component of UG. Colloquium presented at UCLA, Los Angeles, CA.

Smolensky, Paul. 1997. Constraint interaction in generative grammar II: Local conjunction or random rules in Universal Grammar. Handout of talk presented at Hopkins Optimality Theory Workshop/Maryland Mayfest, Baltimore, MD.

Smolensky, Paul. 2005. Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature-domain markedness. In *The harmonic mind: From neural computation to optimality-theoretic grammar.* Vol. 2, *Linguistic and philosophical implications*, by Paul Smolensky and Géraldine Legendre, 590–716. Cambridge, MA: MIT Press.

Wilson, Colin. 2003. Experimental investigation of phonological naturalness. In *WCCFL 22: Proceedings of the 22nd West Coast Conference on Formal Linguistics*, ed. by Gina Garding and Mimu Tsujimura, 533–546. Somerville, MA: Cascadilla Press.

Department of Linguistics and Modern Language Studies
Hong Kong Institute of Education
10 Lo Ping Road
Taipo
New Territories
Hong Kong

ryklai@ied.edu.hk