ASYMMETRIES IN GENERALIZING ALTERNATIONS TO AND FROM INITIAL SYLLABLES

MICHAEL BECKER
University of Massachusetts Amherst

ANDREW NEVINS
University College London

JONATHAN LEVINE
Harvard University

In the English lexicon, laryngeal alternations in the plural (e.g. *leaf* ~ *leaves*) impact monosyllables more than finally stressed polysyllables. This is the opposite of what happens typologically, and would thereby run contrary to the predictions of INITIAL-SYLLABLE FAITHFULNESS. Despite the lexical pattern, in a wug test we found monosyllables to be impacted no more than finally stressed polysyllables were—a 'surfeit of the stimulus' effect, in which speakers fail to learn a statistical generalization present in the lexicon. We present two artificial-grammar experiments in which English speakers indeed manifest a universal bias for protecting monosyllables, and initial syllables more generally. The conclusion, therefore, is that speakers can exhibit spontaneous learning that goes directly against the evidence offered by the ambient language, a result we attribute to formal and substantive biases in phonological acquisition.*

Keywords: initial-syllable protection, laryngeal alternations, English fricative voicing, artificial-grammar learning, underlearning

1. INTRODUCTION. It is now quite generally known that 'underlearning' exists for unnatural linguistic patterns, and that language learners do not demonstrate equal willingness to generalize every statistically robust pattern in their lexicon. For example, focusing on patterns created by morphophonological alternations, Becker et al. 2011 presents results in Turkish in which speakers learned natural phonotactic patterns and did not generalize unnatural patterns, even when both were robustly present in their lexicon. Similar 'surfeit of the stimulus' effects can be found in work such as Zimmer 1969, Zhang & Lai 2008, and Kager & Pater 2012. By contrast, Hayes et al. 2009 discusses results in Hungarian allomorph selection in which unnatural patterns are indeed underlearned, though not completely ignored. In the current article, our point of departure is a particular unnatural pattern found within the lexicon of English, in which monosyllables are impacted by laryngeal alternations (e.g. *knife* ~ *knives*) more than polysyllables are. This pattern would seem to contravene the widespread crosslinguistic generalization that initial syllables are more protected from alternations than noninitial syllables (Trubetzkoy 1939, Steriade 1994, Beckman 1997, 1998, Casali 1998, Alber 2001). We show that the pattern in the English lexicon goes unlearned in a nonce word task, and subsequently demonstrate that in two distinct artificial-language experiments, English speakers in fact prefer the opposite pattern, and asymmetrically protect initial syllables.

The larger contribution of this article is to present evidence that phonological learning operates differently when generalizing FROM initial syllables to other syllables than when generalizing TO initial syllables from other syllables. We demonstrate that generalization in the former direction is much more robust, even when the two patterns are given symmetrical statistical representation in the learning sequence (and in fact, even when the balance is decidedly tipped in the opposite direction, as in the lexicon of English).

* For their valuable comments and discussion, we thank Adam Albright, Lauren Eby Clemens, Peter Graff, William Idsardi, Peter Jurgec, John Kingston, Joe Pater, John McCarthy, Marc van Oostendorp, Anne Pycha, Anne-Michelle Tessier, Marijn van 't Veer, Ruben van de Vijver, Arun Viswanath, Leo Wetzels, Matt Wolf, and the audiences at NELS 40, the LSA 85th annual meeting, and OCP8. Thanks to Parker Crane, Jacob Hoover, and Paul VanKoughnett for their work on different parts of this project. We also thank the editor, associate editor, and two anonymous referees for many helpful comments and suggestions.

We start with a survey of initial-syllable protection effects in a variety of languages and establish the crosslinguistic preference for protecting initial syllables and monosyllabic stems—a particular subclass of initial syllables—from alternations. We then offer results from a survey of the English lexicon, showing that English has a preponderance of alternating monosyllabic items (e.g. *leaf ~ leaves*). This anomaly of the English lexicon is overturned in the following section with a nonce word task ('wug test', Berko 1958), in which the expected results demonstrate that monosyllables are impacted no more than finally stressed polysyllables are. In other words, English speakers seem to ignore the anomalous distribution in the lexicon, and instead adopt a grammar that impacts monosyllables and polysyllables equally.

Next, we present results from an artificial-language experiment that aims to detect a covert bias for the protection of monosyllables in English speakers. We show that English speakers choose to asymmetrically protect monosyllables, confirming the same bias that shapes the typology we have surveyed. We then go beyond monosyllables and present results demonstrating that English speakers protect the initial syllables in polysyllables as well. These findings are interpreted in terms of a universal approach to positionally specific faithfulness, attributing the protection of monosyllables to initial-syllable protection (Trubetzkoy 1939, Steriade 1994, Beckman 1997, 1998, Casali 1998, Alber 2001, Barnes 2006, Becker 2009, Jesney 2009, Becker et al. 2011). We finish with some concluding remarks.

2. Typology: initial syllables are protected from alternations. We begin by surveying the extant typology of size-dependent local morphophonological alternations. We show that when alternations are sensitive to phonological size, they are skewed toward protection of monosyllables. This means that there is an implicational relationship: alternations in monosyllables imply alternations in polysyllables, but not vice versa. In terms of the models proposed by Berwick (1985) and Manzini and Wexler (1987), a language that allows alternations only in polysyllables is a subset language, and serves as the learner's default assumption. In optimality theory, subset relationships are expressed in terms of ranking biases, such as the preference for specific faithfulness over general faithfulness (Smith 2000, Hayes 2004, Tessier 2007). Exposure to alternations in monosyllables will force the learner to adopt a more permissive grammar that allows alternations in both monosyllables and polysyllables. Crucially, there is no grammar that allows productive alternations only in monosyllables, though such a pattern could be listed in the lexicon.

We start with an example of what we mean by a 'local' morphophonological alternation, using the Slovenian adjectival suffix [-ən]. As the examples in 1 show, this affix palatalizes the stem-final consonant. This alternation applies to all nouns that end in [k], as well as to several other consonants (Peter Jurgec, p.c.; see also Toporišič 2000). But the alternation is morphophonological in that it is specific to this affix, as seen from the lack of palatalization in the monomorphemic [ˈpə́kəw] 'hell', and the lack of alternation before a different schwa-initial suffix, as in the diminutive [ˈʧúk-əʦ] 'owl.DIM'.

(1) A regular local morphophonological alternation in Slovenian

BASE NOUN	ADJECTIVE	
ˈbó**k**	ˈbó**ʧ**-ən	'hip'/'lateral'
ˈzná**k**	ˈzná**ʧ**-ən	'sign'/'marked'
baˈɾó**k**	baˈɾó**ʧ**-ən	'Baroque.N/ADJ'
oˈtɾɔ́**k**	oˈtɾɔ̀**ʧ**-ən	'child'/'childish'

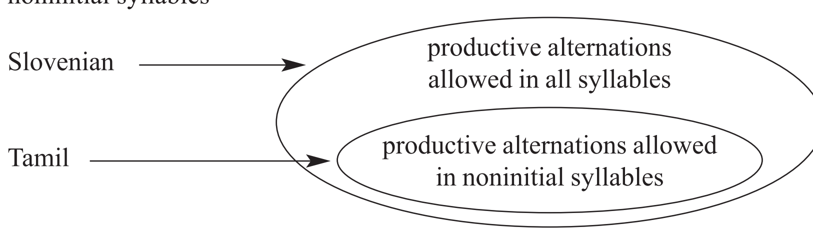
Tamil, by contrast, differs from Slovenian in that MONOSYLLABLES are protected from alternations, as illustrated in 2. Here, [n] becomes [ŋ] before the plural suffix [-ɡə], but [n] is protected from change in the initial syllable (Christdas 1988, Beckman 1997, 1998).

(2) Monosyllables protected from nasal assimilation in Tamil

SINGULAR	PLURAL	
miː**n**	miː**n**-ɡə	'fish'
maːn	maːn-ɡə	'deer'
makə**n**	makə**ŋ**-ɡə	'son'
pajːə**n**	pajːə**ŋ**-ɡə	'boy'

What's missing from the picture is the mirror-image of Tamil, that is, a local morphophonological alternation that impacts stem-final segments in monosyllables but NOT in polysyllables. We schematize the range of known languages with the Venn diagram in 3.

(3) Productive alternations in initial syllables imply productive alternations in noninitial syllables



An even wider range of examples becomes available if we look at alternations that admit lexical exceptions. The best-known example is that of the stem-final laryngeal alternations in Turkish, where the process impacts the majority of polysyllables but only a minority of the monosyllables (Lewis 1967, Hayes 1995, Inkelas & Orgun 1995, Inkelas et al. 1997, Kallestinova 2004, Petrova et al. 2006, Pycha et al. 2007, Becker 2009, Becker et al. 2011). As a representative example, the final stop of the polysyllabic [k^haˈnat^h] 'wing' surfaces as voiced in the possessive [k^hanaˈd-ɨ], but the stop is unchanged in the monosyllabic [ˈk^hat^h] 'floor', where we find possessive [k^haˈt^h-ɨ]. As Becker et al. 2011 has shown, this syllable-count-based asymmetry is an active generalization in the grammar of Turkish speakers, as evidenced by the experimentally verified projection of the pattern from the lexicon onto nonce words.

Two examples of the same type come from French and Brazilian Portuguese, as shown in Becker et al. 2012. In Brazilian Portuguese, the plural induces a change of word-final [w] to [j] only in some lexical items (Huback 2007, Gomes & Manoel 2010), as exemplified in 4. Becker et al. 2012 reports that the alternation impacts only 29% of the monosyllables compared with 88% of the polysyllables, and that in the wug tests conducted, speakers extended this tendency from the lexicon to their treatment of nonce forms, exactly like in Turkish.

(4) Monosyllables protected more than polysyllables in Portuguese

	SINGULAR	PLURAL	
29%	ˈsa**w**	ˈsa**j**s	'salt'
71%	ˈpa**w**	ˈpa**w**s	'stick'
88%	deˈda**w**	deˈda**j**s	'thimble'
12%	kaˈka**w**	kaˈka**w**s	'cocoa'

In both Turkish and Brazilian Portuguese, the alternation is extended from native nouns to loanwords, and in both languages we see the alternation apply more readily to polysyllables. As representative examples, Turkish has the nonalternating monosyllable [ˈtʰypʰ ~ tʰyˈpʰ-y] 'tube' vs. the alternating polysyllable [guˈrupʰ ~ guruˈb-u] 'group'. In Portuguese borrowings from English we find the nonalternating monosyllable [ˈgow ~ ˈgow-s] 'goal' vs. the alternating polysyllable [kokeˈtɛw ~ kokeˈtɛj-s] 'cocktail'. In other words, the size effect is seen in these languages not only as a trend in the lexicon and as an effect of an experimental setting, but also in the natural treatment of new lexical items.

In French, the plural induces a change of word-final [al] and [aj] to [o], as in [mal ~ mo] 'evil SG/PL' and [tʁavaj ~ tʁavo] 'work SG/PL'. Becker et al. 2012 reports that the alternation impacts monosyllables less often than polysyllables, but the evidence in the lexicon is rather modest, since the small number of relevant monosyllables limits the confidence in the size effect. In a nonce word task, however, speakers amplified this pattern, and rated the alternation in polysyllables significantly more acceptable than in monosyllables. In this case, the French speakers are provided with rather weak indication about the distribution of the alternation, yet they generalize exactly like Turkish and Brazilian Portuguese speakers, in preferring to protect nonce monosyllables from the alternation.

Finally, Gouskova and Becker (2012) report that Russian mid vowel deletion (yer deletion) applies more frequently to polysyllables, for example, [buˈgor ~ buˈgr-ov] 'mound (NOM.SG/GEN.PL)', while monosyllables usually resist deletion, for example, [ˈtor ~ ˈtor-ov] 'torus (NOM.SG/GEN.PL)'. This effect is again extended from the lexicon to nonce words.

We propose that in cases in which speakers protect monosyllables, they attribute this protection to the activity of INITIAL-SYLLABLE FAITHFULNESS constraints (Steriade 1994, Beckman 1997, 1998, Casali 1998, Barnes 2006, Becker 2009, Jesney 2009, Becker et al. 2011). In a monosyllable, effecting a change in the word-final segment impacts the WORD-INITIAL syllable, as in the Portuguese [ˈsaw ~ ˈsajs] 'salt(s)', while the initial syllable stays unchanged in a polysyllable such as [deˈdaw ~ deˈdajs] 'thimble(s)'.[1]

However, some researchers have challenged the phonological encoding of initial-syllable faithfulness, and have suggested that size effects are derived from phonetic or psycholinguistic sources. Barnes (2006:184–92) proposes that the source is phonetic duration: initial syllables are phonetically longer in Turkish, and longer segments enjoy greater faithfulness. While we are sympathetic to this line of inquiry, we note that lengthening of initial syllables is not universal. As Barnes reports, initial syllables in English are shortened, not lengthened. The argument against a strictly phonetic basis for protecting monosyllables is made in Becker et al. 2012, which shows that monosyllabicity is the strongest predictor of participants' responses to nonce words in French and Portuguese, leaving no explanatory power left for phonetic duration.

Another strand of research, found in Ussishkin & Wedel 2009 and Stausland Johnsen 2013, suggests that short stems are protected from alternations by psycholinguistic rather than phonetic factors. Specifically, these authors claim that short stems are pro-

[1] In Turkish, the alternating segment does not technically surface in the initial syllable of the word when a vowel-initial suffix is added, as in [ˈkʰatʰ ~ kʰa.ˈtʰ-ɨ], posing an implementational question for formalization in classical optimality theory. This issue is addressed by Jesney (2009), who notes that for a variety of reasons, the initial syllable should refer to the fully faithful parsing of the base [ˈkʰatʰ], not to the winner. This assumes a phonological parsing of the stem as [ˈkʰatʰ] before any affixes are added, along the lines suggested in Wolf 2008.

tected because shorter stems are in denser lexical neighborhoods, on average, than longer stems. Yet Pycha and colleagues (2007) show that in Turkish, such lexicon-based measures have very weak effects, and Becker and colleagues (2011) report that these weak effects are orthogonal to the simple monosyllabicity criterion. Becker et al. 2012 demonstrates that neighborhood density is secondary to monosyllabicity as a predictor of participants' responses to nonce words, and in addition that monosyllables are equally protected regardless of the complexity of their onset; thus the French nonce words [ʁal] and [dal] are just as protected as [stal] and [pnal].

We contend that initial-syllable faithfulness can be grammatically incorporated, beyond its potential functional motivation, and can thereby become an atomic symbolic unit of phonology when it is PHONOLOGIZED (Hyman 1976, 2008). We leave open the possibility, however, that additional functional factors may be identified in the future.

We have focused thus far on local, suffixal morphophonological alternations, where the suffix's impact is on the immediately adjacent segments of the stem. While we have limited specific information on long-distance size-dependent alternations, their profile may turn out to be different. Umlaut, for instance, is known to be attracted to stressed syllables in both German (Wiese 1996, Fanselow & Féry 2002, van de Vijver & Baer-Henney 2011) and Chamorro (Topping 1968, Chung 1983). In the German plural, umlaut impacts only initial syllables—the exact opposite of Turkish. In Chamorro, umlaut seems to impact stressed initial syllables with all umlauting prefixes. Romance metaphony similarly impacts the stressed syllable (Hualde 1989, Walker 2012). Another family of cases about which less is typologically known involves prefixation, which sometimes impacts the left edge of the stem, and thus its initial syllable. Blust (2004) reports that most Austronesian languages with monosyllables prevent nasal substitution on those monosyllables—but substitution is also dispreferred in trisyllables. This is the case in Malay (Delilkan 2005), which protects monosyllables (all of them closed syllables) and trisyllables with an initial closed syllable; thus it is not entirely clear that protection of monosyllables is needed independently from other prosodic restrictions. These questions call for future wug-testing to explore the exact nature of the generalization and its productivity, and we conclude in the interim that while the situation is clear for local suffixation, there is more to be uncovered in the domain of nonlocal alternations and prefixation.

In light of the considerations above, we largely limit our discussion in this article to local, suffixal morphophonological alternations (though see our use of a prefix in the experiment in §7) and leave the very interesting question of nonlocal and prefixal alternations to future work. It may turn out that there are two forces at work, one that protects monosyllables and one that protects initial syllables, and that the two completely overlap in many of the available cases. At present, however, we pursue a more restrictive theory, in which initial syllables are protected, and there is no additional specific protection of monosyllables.

The suffix-induced alternations we have discussed support an asymmetrical view of size-based restrictions on local alternations: monosyllables are protected from alternations, and alternations in monosyllables imply alternations in polysyllables. As for the ultimate source of this effect, we cast doubt on the viability of purely phonetic explanations, and while we leave open the possibility of an eventual psycholinguistic explanation, we adopt here initial-syllable faithfulness as an atomic element of universal grammar that regulates these size effects. It is important to underscore, however, that this article is not about the source of size effects, but rather about their asymmetrical application. Whatever force is at play, it protects monosyllables more than polysyllables and not the other way around. Furthermore, this force protects not just the existing words of a language, but also novel words and words in artificial-grammar tasks.

Interestingly enough, the clearest example of a language with an apparent counter-typological trend along these lines is English, which at first blush seems to furnish a counterexample as a language with alternations that impact monosyllables more strongly than polysyllables. We turn to the facts in the next section.

3. Experiment 1: monosyllables impacted in the english lexicon. In English, final [f] and [θ] alternate with the voiced [v/ð] in some nouns (5a), but not others (5b).

(5) Irregular morphophonological alternations in English

	SINGULAR	PLURAL	
a.	naɪ**f**	naɪ**v**z, *naɪ**f**s	'knife'
	wʊl**f**	wʊl**v**z, *wʊl**f**s	'wolf'
	pæ**θ**	pæ**ð**z, *pæ**θ**s	'path'
	oʊ**θ**	oʊ**ð**z, *oʊ**θ**s	'oath'
b.	stɪ**f**	stɪ**f**s, *stɪ**v**z	'stiff'
	kʌ**f**	kʌ**f**s, *kʌ**v**z	'cuff'
	dɛ**θ**	dɛ**θ**s, *dɛ**ð**z	'death'
	mʌn**θ**	mʌn**θ**s, *mʌn**ð**z	'month'

The phenomenon goes back several centuries and has been commented upon for quite some time (Jespersen 1909, Berko 1958, Hayes 2009), though it is not widely discussed in the generative literature. The phenomenon carries several hallmarks of a semi-productive morphophonological process, as evidenced by its application to a growing set of lexical items, and its partial independence from spelling conventions.

This laryngeal alternation (which we occasionally refer to below as a voicing alternation, recognizing, however, that the opposition in English may in fact involve [spread glottis] and not [voice] (Kager et al. 2007), and thus that it may be a 'deaspiration' process rather than a voicing process) is being increasingly extended to items that formerly did not alternate, especially finally stressed polysyllables. College-age American speakers whom we have informally asked regularly offer the plurals *gira[v]es*, *photogra[v]es*, and *psychopa[ð]s*, whereas older speakers tend to find such forms ungrammatical (effects that we return to in the discussion of experiment 1). Note the orthographic ⟨ff⟩ in *giraffes* and the ⟨ph⟩ in *photographs*, which indicate that these nouns did not alternate historically. Voicing alternations are not always reliably recorded in the spelling even for monosyllables, of course: the plural of *roof* is *roo[v]s* for most Americans, yet the spelling ⟨rooves⟩ is quite uncommon. Finally, we note that the alternations are never orthographically indicated for [θ]-final nouns, yet speakers have clear intuitions about such nouns, as in 5 above.

Second, we note that the set of alternating lexical items has both expanded and contracted over the last few centuries. As Jespersen describes, post-[ɹ] voicing is innovative, as in the plurals of [dwoɹf] 'dwarf', [woɹf] 'wharf', and [skɑɹf] 'scarf' as [dwoɹvz, woɹvz, skɑɹvz] respectively. Conversely, many nouns that had concomitant voicing and vowel-length alternations, as in [stæf] ~ [steɪvz] 'staff', switched to leveled paradigms. Finally, we have reports of massive loss of these alternations in parts of the Midwest (Lauren Eby Clemens, p.c.). All of these changes suggest that grammatical forces are at work, shaping the distribution of the voicing alternations in each speaker's grammar. While the alternations primarily affect Germanic words, alternations may be found in non-Germanic words such as *scarf*, and for many speakers, *giraffe*, *chief*, and *psychopath*, as well.

The English voicing alternations are dependent on morphophonological context: while they affect many nouns in the presence of the plural [z], the identical-sounding genitive suffix does not generally trigger alternations, as in *knife's*, *roof's*, *path's*, and so forth. Many speakers report the blocking of the plural alternations in compounds, as in the simplex plural [buðz] 'booths' vs. the compound [tol-buθs] 'toll-booths' (cf. Kiparsky 1982, Selkirk 1982, et seq.).[2] Voicing alternations may be triggered in denominal verbs, without a clear connection to the plural alternations: for some items, only the noun voices (*knives/to knife*, *leaves/to leaf*, *dwarves/to dwarf*); for others, only the verb voices (*beliefs/to believe*, *proofs/to prove*); and others still have the alternation in both (*halves/to halve*, *lives/to live*, *shelves/to shelve*). There are also doublets such as *to prove* (to establish the validity) vs. *to proof* (to inspect a text for errors). Beyond the plural and the denominal verbs, voicing affects a small number of other derived forms, as in *thief* ~ *thievery*, that may reflect the use of the voiced fricative in all derived forms for a given noun, perhaps a lexical conservatism effect (Steriade 1999).

To summarize, then, we see that while voicing alternations are limited in their morphological and lexical scope, they are subject to strong fluctuations, and are only partly constrained by history or spelling. We take these fluctuations to reflect a grammatical organization of the alternations, an organization that is largely phonological. In order to explore the role of prosodic shape (monosyllabicity and stress) in shaping these patterns and to obtain a realistic sense of the range of acceptable plurals, we surveyed a large number of American English speakers from a range of locations and ages, as explained in the following subsections.

3.1. Materials and methods.

Participants. Participants ($N = 200$) were recruited online using Amazon's Mechanical Turk and were paid \$.30 for their time. Amazon's Mechanical Turk is a web application that provides access to a large number of potential participants for survey-based experiments; see Sprouse 2011 and Schnoebelen & Kuperman 2010 about the use of Mechanical Turk in linguistic research.

Amazon securely stores each participant's name and home address, and associates the participant's work with a unique identification number, which we can utilize to ensure that speakers do not repeat the task. We have used these unique numbers to exclude these participants from similar experiments reported later in this article. Amazon ensures that the participants are located in the United States and are over eighteen years old.

The server logs indicate that our 200 participants took on average thirteen minutes to complete the experiment (range 7–44 minutes, median 12). At the end of the experiment, participants had the option to volunteer demographic information. Year of birth was provided by 152 participants, who reported an average age of thirty-four (range 18–70, median 29). Gender was reported by 127 females and sixty-five males; eight did not say. Country of origin was reported by 180 from the United States, four from another English-speaking country, and two from another non-English-speaking country; fourteen did not say. As for the variety of English spoken, almost everybody named a state or a major city in the US, or said they speak general American English. One hundred and fifteen participants indicated that they were monolingual, thirty-seven indicated some knowledge of Spanish, twenty-six indicated some knowledge of French, and others reported knowledge of a few other languages.

[2] If indeed the form *psychopath* were treated as a pseudo-compound, that is, *psycho-path*, that should serve to discourage the plural *psychopa[ð]s*, yet this alternating plural is quite acceptable (as found below), suggesting that *psychopath* is not analyzed as a compound.

MATERIALS. We extracted all of the [f]-final and [θ]-final words from the CMU dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict) and chose a set of 126 nouns from them, aiming to represent a wide range of prosodic shapes and rhymes. The full list of items, with results by item, is provided in Appendix A. The list contains eighty-two monosyllables (fifty-two [f]-final, thirty [θ]-final), twenty-one polysyllables with a stressed final syllable, or 'iambs' (thirteen [f]-final, eight [θ]-final), and twenty-three polysyllables with an unstressed final syllable, or 'trochees' (twelve [f]-final, eleven [θ]-final).

We note that the CMU lists plural forms for ninety-two of these 126 words. Listed plurals are voiced for twenty-one of sixty-seven monosyllables and for none of the remaining twenty-three polysyllables, a significant difference (Fisher's exact test, $p < 0.01$). This is a first indication that the voicing alternation applies to monosyllables more than polysyllables; the data from our participants validate this observation.

In addition to the [f]- and [θ]-final items, we selected fifty filler items, of which seventeen were monosyllables, fifteen iambs, and eighteen trochees. These included items whose plural is formed by umlaut (e.g. *goose*, *mouse*), infrequent plural suffixes (e.g. *ox*, *appendix*), and items similar to those (e.g. *box*, *human*, *status*).

The items were recorded by a phonetically trained male native speaker of English in his twenties from Maine. The list included each noun in the singular and two plurals, which were presented to the speaker in random order three times. The recording was digitally captured in a sound-attenuated booth. The best token was chosen and converted to mp3 format, and the recordings were not manipulated in any other way, other than normalizing the intensity with Praat (Boersma & Weenink 2011). The speaker regularly lengthened vowels before voiced final fricatives, and pronounced many final fricatives as voiceless, as is common in general American English. Thus, the voicing contrast was often largely borne by a difference in allophonic vowel length.

TASK. The experiment was presented to the participants over the internet, using the web browser of the participant's choice. The web server started by making a random selection of items for each participant, choosing a total of thirty-six items: twenty target items and sixteen fillers.

Item selection was performed by including six monosyllables (three [f]-final and three [θ]-final), four trochees (two [f], two [θ]), and a total of ten iambs: six monomorphemic iambs (four [f], two [θ]), and four polymorphemic iambs (two [f], two [θ]). The fillers were six monosyllables, five trochees, and five iambs, in addition to the training item, *cactus*.

Before the experiment began, participants were reminded that some English plurals are exceptional; they were provided with *ox*, *knife*, and *fish* as examples, and were asked for their preferred plural.

The items were presented orthographically on the screen, with buttons to play two plural forms in random order. Once both plurals were played, a seven-button scale appeared between them, as schematized in 6. This order ensured that the participant heard both plurals before they offered their preference between them.

(6) 🔊 [1] [2] [3] [4] [5] [6] [7] 🔊
(one plural) (other plural)

After the participant pressed one of the seven buttons of the scale, an additional question appeared, asking about the SECOND plural the participant heard. These questions were used to ensure that the participants were listening to the audio material and hearing the voicing distinctions, and asked about vowel quality or voicing, for example, 'did the second plural you heard have the "f" sound of "fan" or the "v" sound of "van"?'. After the participant responded to all thirty-six target items and fillers, they were asked to answer a few demographic questions (year of birth, gender, country of origin, variety of English spoken, and other languages spoken).

3.2. Results. Figure 1 shows the participants' responses by prosodic shape, with monosyllabic items receiving the highest voiced ratings (3.57 on the 1–7 scale), followed by iambs (3.18) and trochees (2.88). Error bars represent 95% confidence intervals. In the following discussion, '7' always refers to the highest acceptability for the voiced plural, and '1' refers to the highest acceptability of the faithful, voiceless plural. For the participants, the ends of the scale were randomly assigned to voiced or voiceless; the raw response was reversed for the trials where the voiceless response was on the '7' side.

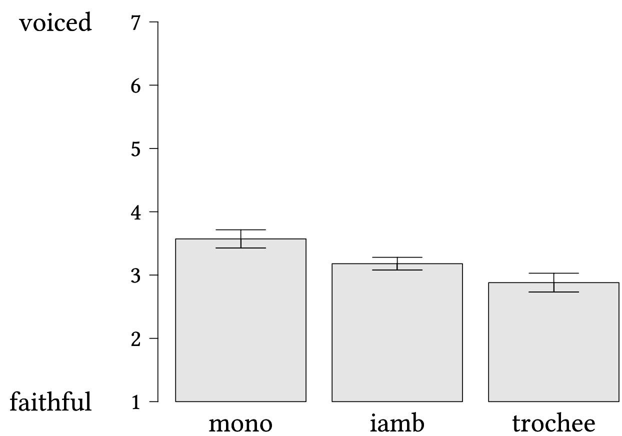


Figure 1. Voicing alternations preferred with monosyllables in real words.

The visual representation of the average responses to each item (plotted with horizontal jitter in Figure 2) shows that only monosyllables such as *elf* and *life* are unanimously voiced (7 on the 1–7 scale), while iambs reach a maximum of 4.67 with *vermouth*, and trochees are even lower with 4.03 for *Behemoth*.

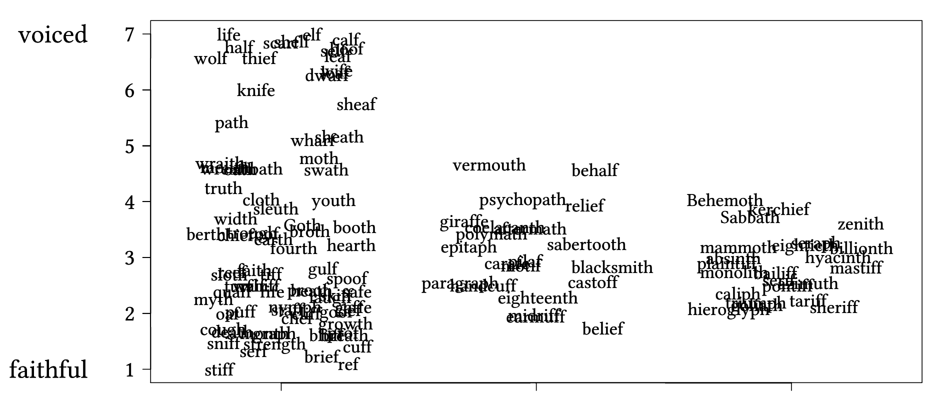


Figure 2. Voicing alternation ratings for real words, by item.

These findings indicate that all of the items that are strongly impacted by the voicing alternation are monosyllabic. None of the polysyllables we tested reached a voicing score of 5, leaving the entire top third of the range to the monosyllables.

A comparison of particular interest, to which we return in §4, is the difference between monosyllables and iambs, in which stress is kept consistently on the alternating syllable. This is a positional effect, as the alternations in iambs occur in a noninitial syllable.

The difference between monosyllables and iambs on the one hand and trochees on the other is a stress effect: in the trochees, the voicing alternation causes the appearance of a voiced fricative in an unstressed syllable, where the contrast between voiced and voiceless fricatives is less robust (see Giavazzi 2010 for a recent survey of such effects). This is especially true in English, where the contrast is mostly cued by vowel length, which in turn is diminished in an unstressed syllable.

Voiced responses were preferred following the long vowels [æ,[3] ɑ/ɔ,[4] i, u, eɪ, oʊ, aɪ, aʊ] relative to the short vowels [ɛ, ɚ, ɪ, ʊ, ʌ] (3.41 vs. 2.66 on the 1–7 scale). This vowel length difference was strong for the monosyllables (3.94 vs. 2.83) and iambs (3.34 vs. 2.29), but not for the trochees (2.88 vs. 2.89). The preference for voiced consonants following a long vowel is well attested crosslinguistically, and is also mirrored in a similar experiment on voicing alternations in Dutch (Ernestus & Baayen 2003; see a discussion of their vowel effect in Becker et al. 2011).

Morphological complexity had the expected effect, with morphologically complex forms less acceptable with voicing alternations, but the effect was rather small (3.33 vs. 2.94).

Since the items in this experiment are real words, we checked for the effect of token frequency, plotted in Figure 3 with a *lowess()* trend line. We used the frequencies supplied by the Microsoft Web N-gram Services (based on the Bing search engine). Token frequency had a weak, nonsignificant positive correlation with the responses (Spearman's rank correlation, $\rho = .06$, $p > 0.1$). This is in line with findings in Bybee 1995, Albright & Hayes 2002, Hay et al. 2004, and Becker et al. 2011, which propose that frequency can affect the behavior of individual items, but that overall trends are generally sensitive to the types in the lexicon, rather than being directly affected by token frequencies.[5] According to Moder (1992), learners generalize less from frequent items, and it is indeed the case that in our materials, monosyllables have higher token frequency than polysyllables (two sample *t*-test, $t(90.6) = 5.8$, $p < 0.0001$).

To assess the statistical strength of these effects, we employed a mixed-effects regression model using the *lmer()* function from the *lme4* package (Bates & Maechler 2009) in R (R Development Core Team 2011). The model included the following predictors: *long vowel*, a binary predictor that encoded vowel length as explained above; *shape*, a three-level unordered factor that distinguished monosyllables, iambs, and trochees; *morphological complexity*; *place* (f vs. θ); and *token frequency*. The effects that reached statistical significance are reported in Table 1.[6]

[3] While [æ] is historically short, it is phonetically one of the longest vowels of American English, as noted, for example, in Lisker 1974. The remaining vowels in the list are uncontroversially long.

[4] The speaker who recorded the stimuli did not have the *cot/caught* distinction.

[5] At the right edge of Fig. 3, it is clear that the most frequent items tend to be closer to the extremes of the voicing continuum, while less frequent items are closer to the middle. More frequent items are uttered more often in the plural, and can become entrenched as either clearly voiceless or clearly voiced, while the behavior of the infrequent items is less fully determined. This effect is statistically significant, with token frequency being predictable from the square of the voicing score. This effect is largely tangential to the question we ask here, which is how the speaker generalizes from known items to unknown items.

[6] The model in Table 1 was computed in the following steps. We started with a base model that had *item* and *participant* as random effects, and no fixed effects. We then added *long vowel*, an addition that made a

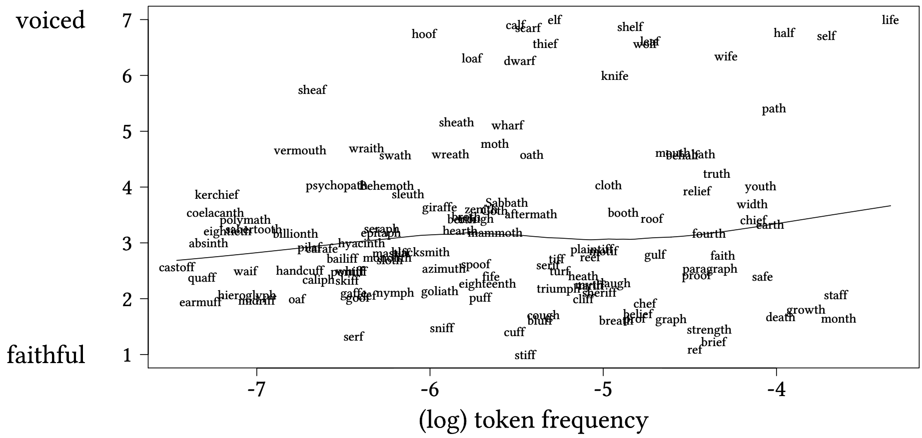


FIGURE 3. Weak, nonsignificant token-frequency effect in real words.

	β	SE(β)	t	p-value
(intercept)	3.19	0.12	25.53	
mono vs. iamb	–0.28	0.12	–2.38	< 0.05
mono & iamb vs. trochee	–0.20	0.07	2.76	< 0.01
long vowel	0.26	0.09	2.76	< 0.01

TABLE 1. Lexicon model.[7]

The model in Table 1 reflects the fact that having a long vowel is conducive to more fricative voicing (positive β), while being iambic or trochaic is less conducive to voicing (negative β for both).

To assess the effect of age, we ran a model on the data from the 152 participants who supplied their year of birth. Adding *age* as a predictor did not improve the model. A simple by-participant analysis showed that the correlation between average voicing scores and age was strongest for iambs, but did not reach significance (Spearman's rank correlation, $\rho = -.14$, $p > 0.1$). We conjecture that the age effect we anecdotally noted among our university consultants may be limited by class or level of education, and thus did not manifest itself sufficiently strongly in a more diverse participant pool.

To summarize, we found two factors that affect the voicing of existing [f]-final and [θ]-final nouns: the length of the noun's final vowel, with long vowels conducive to

significant improvement to the model (ANOVA model comparison, $\chi^2(1) = 7.0$, $p < 0.01$). Next, we added *shape*, which made another significant improvement to the model that contained *long vowel* (ANOVA model comparison, $\chi^2(2) = 6.86$, $p < 0.05$). The interaction of *long vowel* and *shape* did not improve the model significantly. Other predictors, such as morphological complexity, place (f vs. θ), and token frequency, made no significant improvement to the model. To reduce the correlations between the predictors in the model, we first normalized *long vowel* using R's *scale()* function. Then, we helmert-coded *shape* as two binary predictors, one that contrasts monosyllables with iambs, and one that contrasts monosyllables and iambs with trochees, and normalized each predictor. Finally, we made a fully crossed model, which had *long vowel* and *shape* as random slopes for both *participant* and *item*. This is the model we report in Table 1. The model has low collinearity measures (VIF ≤ 1.02, $\kappa = 1.27$), calculated using *mer-utils*, by Austin Frank (https://hlplab.wordpress.com/2011/02/24/diagnosing-collinearity-in-lme4/).

[7] Exact p-values cannot be obtained with *lme4* or *pvals.fnc()*, due to current disagreement in the community about the best way to calculate them. We obtained p-values by taking the three predictors out of the model, one at a time, and testing their improvement to the superset model; this method is widely accepted as reliable and impervious to collinearity. A recent discussion of these matters can be found at http://hlplab.wordpress.com/2010/05/10/mini-womm/.

significantly more voicing, and the shape of the noun, with monosyllables conducive to significantly more voicing than iambs, and monosyllables and iambs conducive to significantly more voicing than trochees. The effects of morphological complexity, place (f vs. θ), token frequency, and age were not strong enough to reach statistical significance.

3.3. Discussion. This survey provides a good starting point for a study of English voicing alternations. We asked a large number of speakers about most of the [f/θ]-final nouns in English, and established that only monosyllables score highly on the voicing scale. Out of the eighty-two monosyllabic nouns tested, nineteen scored over 5 (on a scale of 1–7), while none of the polysyllabic nouns reached this rating level.

4. Experiment 2: no monosyllabicity effect in wugs. The experimentally conducted survey reported in §3 demonstrates that in the English lexicon, monosyllables are impacted by voicing alternations more strongly than polysyllables, a finding that confirms dictionary data. This finding is surprising given the typological background discussed in §2: monosyllables quite generally enjoy greater protection from alternations. To gain insight into the grammar that speakers use to regulate alternations and the productivity of this pattern, we contend that it is necessary to go beyond the existing words of the language.[8]

4.1. Materials and methods.

Participants. Participants (*N* = 200) were recruited through Amazon's Mechanical Turk and did not participate in the previous experiment. Participants were paid $.30 for their time. The server logs indicate that these participants took on average fourteen minutes to complete the experiment (range 6–54 minutes, median 12). Participants had the option to volunteer demographic information. Year of birth was provided by 140 participants, who reported an average age of thirty-one (range 18–62, median 27). Gender was reported by 126 females and sixty-two males; twelve did not say. Country of origin was reported by 173 from the United States and one from Canada; twenty-six did not say. As for the variety of English spoken, almost everybody named a state or a major city in the US, or said they speak general American English. One hundred and eighteen participants indicated that they were monolingual, forty-one indicated some knowledge of Spanish, twenty-seven indicated some knowledge of French, and others reported knowledge of a few other languages.

Materials. To test the effect of vowel length and prosodic shape on fricative voicing in the plural of novel nouns, we created 132 target nonce words ('wugs'). We made these by crossing two places of articulation (f, θ) * three prosodic shapes (monosyllable, iamb, trochee) * eleven vowel categories (æ, aɪ, ɑ/ɔ,[9] eɪ, ɛ, ɚ, i, ɪ, oʊ, u/ʊ,[10] ʌ), and then constructed two words in each of these combinations, selecting consonants to match general English frequencies, allow reasonable well-formedness, and keep sufficient distance from real fricative-final nouns. The full list of nonce items, with results

[8] To our knowledge, this may be the first study to wug-test the English fricative voicing alternations in several decades, and the first to include polysyllables. Berko (1958) included the nonce item ⟨heaf⟩ in her classical experiment, eliciting 42% voiced responses from adults, and 3% voiced responses from children. Graves and Koziol (1971) report no voiced responses in spelling from children in grades 1–3 for the items ⟨plif⟩, ⟨heaf⟩, ⟨rife⟩, ⟨truf⟩.

[9] The speaker who recorded the stimuli did not have the *cot/caught* distinction.

[10] We made half as many items for [ʊ] and [u] as we did for the other vowels, to better represent the rarity and marginal status of [ʊ], considering [u] to be long and [ʊ] short.

by item, is given in Appendix B. In addition, we created thirty filler nonce items (thirteen monosyllables, five iambs, twelve trochees), which were modeled on the fillers in experiment 1.

The items were recorded by a phonetically trained male native speaker of English in his thirties from California, following the same procedure as in §3.

TASK. The experiment was presented to the participants over the internet, closely matching the procedure described in §3.1. The web server started by making a random selection of items for each participant, choosing a total of thirty-six items: eighteen target items and eighteen fillers. Six monosyllables (three [f], three [θ]), six trochees (three [f], three [θ]), six iambs (three [f], three [θ]), and the eighteen fillers were chosen randomly. Before the experiment began, the participant was reminded that some English plurals are exceptional, and was provided with *ox*, *knife*, and *phenomenon* as examples. The participants were asked to rate made-up words of English, using the nonce *pharon* as the example word. The rest of the experiment proceeded as described in §3.1, asking for ratings on a 1–7 scale.

4.2. RESULTS. Figure 4 shows the participants' responses in terms of size: iambs receive the highest voicing ratings (4.87 on the 1–7 scale), followed closely by monosyllables (4.38). This is the OPPOSITE of the results in the lexicon study, where monosyllables received the highest ratings for voiced plurals. While the difference between monosyllables and iambs was not significant, trochees came in significantly lower (4.06), as they did in the lexicon study. As in experiment 1, '7' denotes the highest acceptability of the voiced plural, and '1' the highest acceptability of the faithful plural; for the participants, the extremes of the scales were randomly assigned to 1 or 7 for each item.

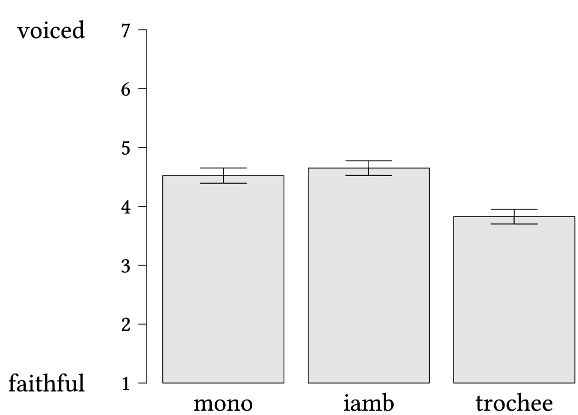


FIGURE 4. Nonce iambs and monosyllables alternate more than trochees.

The same pattern is observed when looking at the average responses by item in Figure 5. Note that unlike real items, most of which have relatively established plural forms, nonce words typically do not elicit extreme responses. We take this to be a validation of the statistical model in Table 1, which predicts highly reliable but numerically small differences between items based on prosodic shape and vowel length.

The results again demonstrate a difference between monosyllables and trochees, the same stress effect observed in the lexicon. Voiced plurals are judged worse in unstressed syllables, both in the lexicon and in nonce words. Where stress is held constant

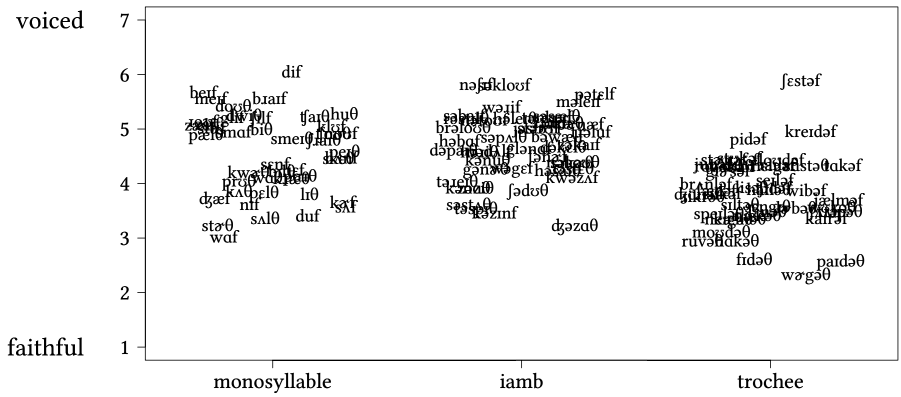


FIGURE 5. Voicing alternation ratings for nonce words, by item.

in monosyllables and iambs, however, the positional effect that was seen in the lexicon is gone: nonce monosyllables are not more strongly impacted by the alternation than nonce iambs.

The statistical analysis employed a mixed-effects regression model with the *lmer()* function, as in §3. As predictors, we used *long vowel*, *shape*, and *place*, defined as in §3. The effects that reached significance are reported in Table 2.[11]

The model in Table 2 indicates that having a long vowel is conducive to more voicing (positive β), and being trochaic is conducive to less voicing (negative β)—both effects also seen in the lexicon study. The effect of vowel length is strongest (.16) for monosyllables, and weaker for iambs (.16 − .09 = .07) and for trochees (.16 − .11 = .05). Being [θ]-final is also conducive to less voicing, even though this effect did not reach significance in the lexicon model. Most importantly, however, there is a trend toward MORE voicing in iambs than in monosyllables, unlike what we have seen in the lexicon.

	β	SE(β)	t	p-value
(intercept)	4.34	0.08	54.81	
mono vs. iamb	0.05	0.05	0.85	> 0.1
mono & iamb vs. trochee	–0.36	0.05	–7.14	< 0.0001
long vowel	0.16	0.05	3.07	< 0.005
place = θ	–0.29	0.05	–6.28	< 0.0001
mono vs. iamb:long	–0.09	0.05	–1.65	> 0.1
mono & iamb vs. trochee:long	–0.11	0.05	–2.04	< 0.05

TABLE 2. Model of fricative voicing in nonce words.

[11] The model in Table 2 is the result of the following steps. We started with a base model that had *item* and *participant* as random effects, and no fixed effects. We then added *shape*, an addition that made a significant improvement to the model (ANOVA model comparison, $\chi^2(2) = 35.3$, $p < 0.0001$). Next, we added *long*, which made another significant improvement to the model that contained *shape* (ANOVA model comparison, $\chi^2(1) = 8.04$, $p < 0.01$). We then added *place*, making another improvement to the model (ANOVA model comparison, $\chi^2(1) = 23.7$, $p < 0.0001$). The interaction of *shape* and *long* made another significant improvement to the model (ANOVA model comparison, $\chi^2(2) = 6.87$, $p < 0.05$), but no other interactions did. We reduced the correlations between the predictors in the model as explained in §3. Finally, the random effects of *item* and *participant* were each augmented with random slopes for the two *shape* predictors, *long*, and their interaction, and this is the model we report in Table 2. The fully crossed model, that is, the one including random slopes for *place*, did not converge (i.e. the magnitude of the effects could not be established, so no model was produced). This model has low collinearity measures (VIF ≤ 1.17, κ = 1.09). We calculated p-values as explained in §3.

The overall acceptability of the voiced plurals is higher in this experiment (4.34 here, 3.19 in experiment 1). While the experiments differ in several respects, we conjecture that the effect is due to the difference in fillers: the fillers in experiment 1 were frequent irregular plurals, and thus highly acceptable. In experiment 2, the fillers were nonce versions of these minority patterns, and thus highly unacceptable. Supposing that some participants had an implicit tendency to distribute their responses toward the alternating end of the scale, in experiment 2, they only had a chance to do so with the voiced plurals, and not with the fillers. In other words, the fricative-final items would seem much more plausible in the context of the implausible fillers, and vice versa.

To assess the effect of age, we ran a model on the data from the 140 participants who supplied their year of birth. Adding *age* as a predictor did not improve the model. A simple by-participant analysis showed a significant and negative correlation between age and voicing ratings, but this correlation was comparable for all three shapes (Spearman's rank correlation, for monosyllables: $\rho = -.20$, $p < 0.05$, iambs: $\rho = -.21$, $p < 0.05$, and trochees: $\rho = -.19$, $p < 0.05$).

To summarize, we found three factors that affect the voicing of novel [f]-final and [θ]-final nouns: the length of the noun's final vowel, with long vowels conducive to significantly more voicing; the place of the final consonant, with [θ] being conducive to significantly less voicing than [f]; and the shape of the noun, with monosyllables and iambs conducive to significantly more voicing than trochees. The difference between monosyllables and iambs trended in favor of the iambs, but did not reach significance, strongly contrasting with the effect found in the real words of English, where iambs were significantly LESS conducive to voicing than monosyllables.

4.3. DISCUSSION. The goal of this experiment was to assess the generalizability of the propensity of monosyllables to alternate relative to iambs in English fricative voicing. In the real words of English, monosyllables are impacted significantly more than iambs, while in the nonce word experiment, monosyllables are trending in the other way, demonstrating more protection than iambs. We take this to mean that the pattern in the English lexicon is underlearned: speakers do not encode the generalization that monosyllables can be impacted more than iambs. As Hayes (2009:195) says, 'it [is] likely that *knives* and similar forms are memorized', in other words, encoded as listed exceptions, somewhat akin to the dual-route model of Prasada and Pinker (1993). Learners are biased to protect monosyllables, and this bias shapes the generalizations that they can extract from the real words they know. This bias against learning unnatural generalizations is what Becker and colleagues (2011) call the SURFEIT OF THE STIMULUS, an effect we demonstrated here with over 100 stimuli and 200 participants (and thereby embraced Hayes and colleagues' call for larger data sets in the investigation of such effects).

The underlearning of this countertypological trend contrasts in this case with the two natural, typologically supported generalizations that the participants did extend from their lexicon onto the nonce words. The first is the effect with long vowels, which are conducive to more alternations in the lexicon, which was productively extended to nonce words. The second concerns the unstressed position, which is conducive to fewer alternations both in the lexicon and in the nonce words. In short, we document the underextension of unnatural patterns alongside the observable readiness to extend patterns that do have a basis in phonological typology.

5. EXPERIMENT 3: ENGLISH SPEAKERS PROTECT MONOSYLLABLES IN AN ARTIFICIAL GRAMMAR. The preceding sections have established that in the English lexicon, monosyllables are impacted by alternations more than polysyllables, but that this anomaly is absent from the treatment of nonce words. The goal of the present experiment is to investigate whether English speakers are biased to protect monosyllables in the setting of an artificial-language learning experiment with different types of alternations.

Our goal was specifically to investigate whether alternations in monosyllables imply alternations in polysyllables, but not the other way around. Artificial-language experiments have been successfully used to show implicational relationships in phonology, most notably by Wilson (2006) and Moreton (2008) to investigate analytic bias; see Moreton & Pater 2011 for a recent review of this literature.

In their real language, English speakers regulate voicing alternations in the plural on [f, θ]. In this artificial language, we maintained plural as the morphological category, but changed the plural suffix to [-ni], and the set of alternating consonants to [p, t, k].

5.1. MATERIALS AND METHODS.

PARTICIPANTS. The participants ($N = 100$) were recruited through Amazon's Mechanical Turk and did not participate in any of the previous experiments. Fifty were assigned to the monosyllabic training group and fifty were assigned to the iambic training group. They were paid $.50 for their time. The server logs indicate that these participants took on average ten minutes to complete the experiment (range 5–25 minutes, median 10).

Participants had the option to volunteer demographic information. Year of birth was provided by all participants, who reported an average age of thirty-one (range 18–62, median 28). There was no significant age difference between the groups (difference in mean < 1 year, two-sample t-test, $t(97.95) = 0.13$, $p > 0.1$). Gender was reported by fifty-three females and forty-seven males. There was no significant gender difference between the groups (monosyllabic training: twenty-two females, iambic training: thirty-one females, Fisher's exact test, $p > 0.1$). Ninety-five participants said they grew up speaking English, three said they did not, and two did not say. As for the variety of English spoken, almost everybody named a state or a major city in the US, or said they speak general American English. Forty-four participants indicated that they were monolingual, twenty-seven indicated some knowledge of Spanish, nine indicated some knowledge of French, and others reported knowledge of a few other languages.

MATERIALS. We constructed sixty stop-final target items and forty-four sonorant-final fillers. The sixty target items, listed in Appendix C, were made by crossing three final consonants (p, t, k) * two shapes (monosyllable, iamb), with ten items in each category created by filling in stressed vowels from the set [ɑ, e, i, o, u, aɪ] and onsets from all the consonants of English except for the interdentals and glides. In the iambs, the initial unstressed vowel was consistently schwa. The fillers were twenty-two monosyllables and twenty-two iambs, with final consonants drawn from the set [l, ɹ, m, ŋ] and nonfinal segments drawn from the same vowels and consonants that were used for the target items.

The items were recorded by a phonetically trained male native speaker of English in his twenties from Pennsylvania. The list included each noun in the 'singular' (i.e. the bare form) and two 'plurals' (one made by simple suffixation of [ni], and one by voicing of the stem-final stop and suffixation of [ni]). The recording was conducted using the same procedure as in §3 and §4.

To assign the experimental items plausible meanings as concrete nouns, we collected 100 pairs of pictures of easily recognizable objects (e.g. *airplane*, *anteater*, *apple*) using Google's image search with the 'line drawing' option. The 'singular' was represented by one of these pictures, and the 'plural' was represented by showing five copies of the other picture arranged in two rows. A sample singular-plural pair is shown in Figure 6.

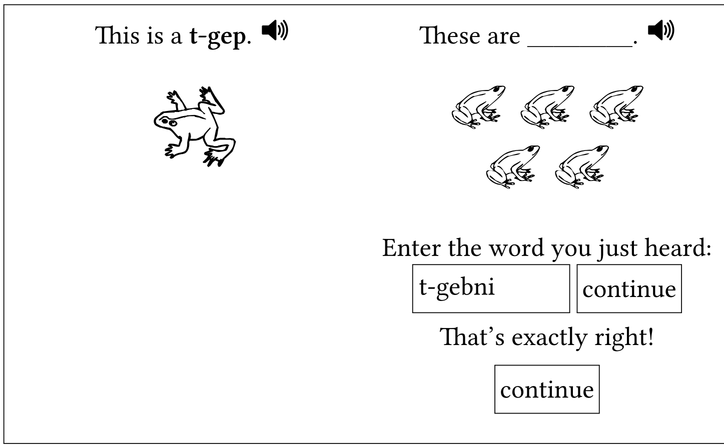


FIGURE 6. Screenshot from experiment 3 showing training with feedback.

TASK. The experiment was presented over the internet as an artificial-grammar learning game. In the beginning of the experiment, the web server randomly assigned each participant to one of two training groups: monosyllabic or iambic.

The training phase consisted of fifteen trials, with ten randomly chosen stop-final items (four [p]-final, three [t]-final, three [k]-final) and five randomly chosen sonorant-final items for each participant (see Table 3). Participants in the monosyllabic training group received ten monosyllabic stop-final nouns, two monosyllabic sonorant-final nouns, and three iambic sonorant-final nouns. Participants in the iambic training group received the exact opposite: ten iambic stop-final nouns, two iambic sonorant-final nouns, and three monosyllabic sonorant-final nouns. Each item was randomly assigned to a pair of pictures as its meaning.

	MONOSYLLABIC TRAINING		IAMBIC TRAINING	
TRAINING	10 STOP-FINAL MONOS		10 STOP-FINAL IAMBS	
	ˈmip	ˈmibni	təˈgep	təˈgebni
	ˈstut	ˈstudni	gəˈʃut	gəˈʃudni
	5 SONORANT-FINALS		5 SONORANT-FINALS	
	ˈmuŋ	ˈmuŋni	ˈmuŋ	ˈmuŋni
	nəˈʤol	nəˈʤolni	nəˈʤol	nəˈʤolni
TESTING	10 STOP-FINAL MONOS		10 STOP-FINAL MONOS	
	ˈgaɪp	—	ˈgaɪp	—
	ˈklet	—	ˈklet	—
	10 STOP-FINAL IAMBS		10 STOP-FINAL IAMBS	
	fəˈʧop	—	fəˈʧop	—
	bəˈgit	—	bəˈgit	—
	10 SONORANT-FINALS		10 SONORANT-FINALS	
	ˈpleɹ	—	ˈpleɹ	—
	ʒəˈtaɪm	—	ʒəˈtaɪm	—

TABLE 3. Artificial-grammar setup for experiment 3.

The training phase began with participants hearing the singular and seeing it written with English-like spelling, and then hearing and seeing the plural with the same English-like spelling. After seven such trials, the next eight trials showed and played the singular, but the plural was only played and not written, and the participant was asked to type it in. Once typed in, participants were shown the intended spelling and an indication of whether their spelling matched the intended one. In this phase, plurals always contained a voicing alternation, for example, [ˈmi**p** ~ ˈmi**b**ni], which was reflected in the proposed spelling.

The testing phase consisted of thirty trials. The server randomly chose twenty stop-final items: ten stop-final items of the shape that was seen before (monosyllabic for the monosyllabic training group and iambic for the iambic training group), and ten stop-final items of the shape on which the participant had not been trained. Of the ten stop-final items of the shape on which they had been trained, four had actually been seen before in the training phase, and six were completely new. The fillers were ten sonorant-final nouns, five of each shape.

In the testing phase, participants first heard the singular and saw it written with English-like spelling, and they were asked to supply the plural. No feedback was given at this phase. After fifteen such trials, the next fifteen trials only played the singular; it was not shown in written form. Participants were asked to supply the plural based on their ability to hear the singular, which thus ensured that they listened to the audio materials.

Throughout the task, the vowels [ɑ e i o u aɪ ə] were spelled ⟨a e i o u ai –⟩, and the consonants [k ʒ] were spelled ⟨k zh⟩. The spelling of schwa as '–' was introduced with the example 'puh-lease', and our participants accurately learned to use this convention.

5.2. Results. In order to analyze the plurals provided as responses for stop-final singulars, we looked at responses that ended in the correct plural suffix (i.e. the spelling ⟨ni⟩, which was produced in 97% of the responses), and then identified the grapheme that preceded this suffix. The graphemes ⟨p⟩, ⟨b⟩, ⟨t⟩, ⟨d⟩, ⟨k⟩, ⟨g⟩ were produced in 94% of all responses, and those were considered valid responses to stop-final items; other responses were considered invalid. The graphemes ⟨b⟩, ⟨d⟩, ⟨g⟩ were considered voiced responses, and ⟨p⟩, ⟨t⟩, ⟨k⟩ voiceless. Each response was categorized as to whether it matched the place of articulation of the singular's final stop; for example, either ⟨p⟩ or ⟨b⟩ in the plural were considered to match a ⟨p⟩ in the singular (84% of all responses).

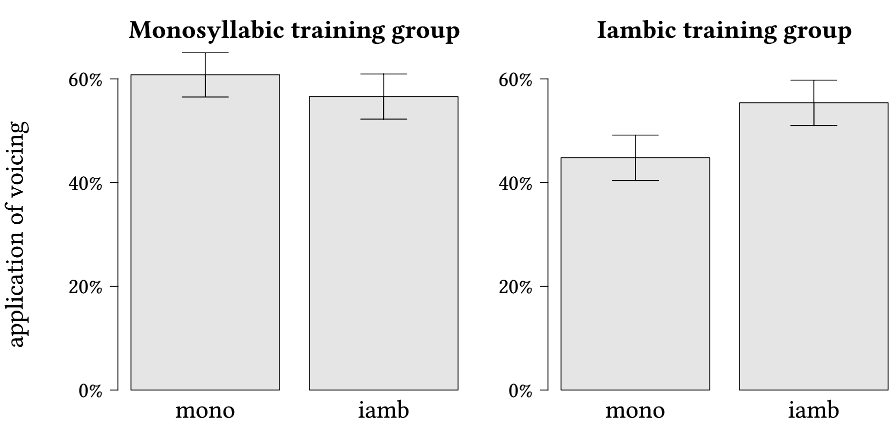


Figure 7. English speakers generalize cautiously to monosyllables, experiment 3.

In the monosyllabic training group, voiced responses were provided for monosyllabic items more than for iambs (61% and 57%, respectively, a small but significant difference). In the iambic training group, voiced responses were provided significantly more often for iambic items than for monosyllabic items (55% vs. 45%, respectively, a larger significant difference), as seen in Figure 7. In both groups, the shape used in the training (monosyllables with monosyllabic training and iambs with iambic training) received voiced responses at essentially the same rate (61% and 55%, respectively, an insignificant difference), suggesting that both groups observed the alternation equally well.

The statistical analysis was performed with an *lmer* mixed-effects logistic regression model, using the following two predictors and their interaction: *group*, a binary factor that distinguished monosyllabic training from polysyllabic training, and *untrained*, a binary factor that coded the prosodic shape that participants did not see in the training, that is, iambs for monosyllabic training and monosyllables for iambic training. The resulting model is reported in Table 4.[12]

	β	SE(β)	t	p-value
(intercept)	0.49	0.31	1.59	
group	–0.28	0.31	–0.91	> 0.1
untrained	–0.30	0.07	–4.15	< 0.0001
group:untrained	–0.21	0.08	–2.78	< 0.01

TABLE 4. Artificial-grammar model for experiment 3.

The analysis reveals no difference between the two groups in their overall propensity to offer voiced responses (no significant main effect for *group*). Furthermore, participants in both groups apply the voicing alternation significantly less to the shape they were not trained on—that is, they extend the alternation cautiously. Of most interest here is the interaction: participants were significantly more reluctant to extend the alternation in the iambic training group than they were in the monosyllabic training group. This is what we expected based on an analytic bias against alternations in initial syllables: with monosyllables being protected by default (as in the subset grammar), participants are reluctant to impact them if they only see the alternation impacting iambs.

Finally, we note that a smaller group of twenty-two Harvard students participated in a pilot for this experiment, with the same materials and similar methods as we describe above, but were asked to speak their responses rather than type them. This much smaller group provided responses that are qualitatively identical to the results we obtained from the Mechanical Turk participants. With this cross-validation in hand, we preferred Mechanical Turk's broader participant pool for its greater coverage in geography, age, and other sociolinguistic factors.

5.3. DISCUSSION. In the training phase, participants were exposed to a fragment of a 'language', only seeing alternations impacting monosyllables or only seeing alternations impacting iambs. We then tested the expectations they formed by testing them on the shape they had not seen. The experimental setup in both groups was symmetric, but the results were asymmetric.

The participants in the monosyllabic training group were trained on a language fragment that resembles real English, where the monosyllables alternate. Recall that for many speakers of English, voicing alternations only impact monosyllables in the listed items in the lexicon. The participants in the iambic training group were trained on a lan-

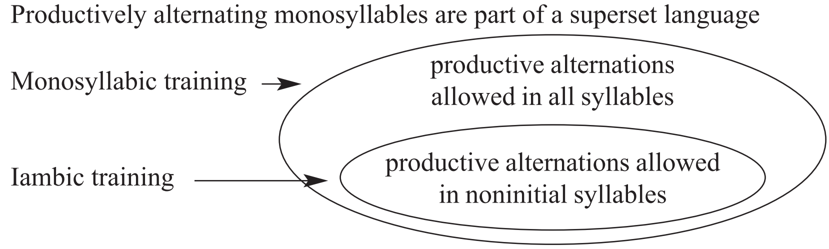
[12] The model in Table 4 is a fully crossed model that had *group*, *untrained*, and their interaction as fixed effects, and the same predictors as random slopes for *item* and *participant*. The model we report in Table 4 uses all 2,000 data points (twenty stop-final stimuli * 100 participants); the models that use only valid responses or only place-matching responses are virtually identical. The model has low collinearity measures (VIF $\leq$ 1.13, $\kappa = 1.0002$).

guage fragment that resembles Turkish, French, or Brazilian Portuguese (see §2), where voicing alternations mostly impact polysyllables.

Participants in both groups were cautious in applying the alternation to the shape they had not seen. Interestingly, however, participants in the iambic training group showed significantly more caution, protecting monosyllables. In other words, although they are English speakers, the participants followed the typologically common pattern, and not the English-specific pattern. This is a clear case of the poverty of the stimulus: in the absence of information about whether the alternation affected monosyllables, speakers deployed an internal bias to protect them. They did not assume that they can freely extend the alternation from polysyllables to monosyllables.

This bias is illustrated with the Venn diagram in 7. Since alternations in initial syllables automatically include alternations in noninitial syllables, monosyllabic training licenses the inference to alternations in all nouns, but iambic training does not license the inference to alternations in monosyllables.

(7) Productively alternating monosyllables are part of a superset language



We note that all of the stimuli had final stress, which placed the alternating stop in the coda of the stressed syllable, and thereby kept the stress constant. The difference between the monosyllables and the iambs is that only in monosyllables does the alternation impact the word-initial syllable.

We conclude that given a chance, English speakers underextend the license that the English lexicon shows for more alternation in initial syllables, and instead prefer the typologically attested protection of initial syllables. The participants went directly against the evidence from their ambient language, showing the spontaneous emergence of an analytic bias.

6. Experiment 4a: initial syllables also protected in polysyllables. The three experiments described in §3, §4, and §5 contrasted monosyllables with polysyllables, and jointly demonstrated an asymmetrical preference to protect monosyllables from alternation. We claimed that the asymmetry is best understood in terms of initial-syllable faithfulness, which can limit alternations in monosyllables but not in polysyllables. These experiments, however, did not test initial-syllable faithfulness in polysyllabic items. To strengthen the connection between the monosyllabicity effect and initial-syllable faithfulness, we conducted an experiment with symmetric training in the groups similar to experiment 3, but in a language where all stems are disyllabic, either trochaic or iambic, and in which nonlow vowels switch their backness in the plural,[13] for example, [ˈzumәp ~ ˈzimәp] and [sәˈfup ~ sәˈfip], loosely inspired by German umlaut.[14] The alternation always occurred in the stressed syllable, and what changes is the position of the alternation: initial or noninitial. Note that we had to switch from a local process,

[13] We limited ourselves to changes that keep vowel height constant, as these have a limited effect on duration. We did not use the pair [ɑ ~ æ], since for many speakers, [æ] is not as low as [ɑ].

[14] English umlaut, which always impacts the initial syllable, is limited to the seven nouns *man*, *goose*, *tooth*, *foot*, *mouse*, *louse*, and *woman* (with some dispute about whether the latter is in the same class). German umlaut is more productive (Wiese 1996, Fanselow & Féry 2002, van de Vijver & Baer-Henney 2011). In

where the affix causes a change on immediately adjacent segments, to a nonlocal process, where stem-internal segments are impacted; there is no other way for a process to impact both potential positions in a set of disyllables.

6.1. Materials and methods.

Participants. The participants ($N = 100$) were recruited through Amazon's Mechanical Turk and had not participated in any of the previous experiments. Fifty were assigned to the trochaic training group and fifty were assigned to the iambic training group. They were paid $.50 for their time. The server logs indicate that these participants took on average ten minutes to complete the experiment (range 5–29 minutes, median 9).

Participants had the option to volunteer demographic information. Year of birth was provided by all participants, who reported an average age of twenty-nine (range 18–60, median 27). There was no significant age difference between the groups (difference in mean ~2 years, two-sample *t*-test, $t(85.93) = 1.13$, $p > 0.1$). Gender was reported by sixty-seven females and thirty-three males. There was no significant gender difference between the groups (trochaic training: thirty-three females, iambic training: thirty-four females, Fisher's exact test, $p > 0.1$). Ninety-six participants said they grew up speaking English, three said they did not, and one did not say. As for the variety of English spoken, almost everybody named a state or a major city in the US, or said they speak general American English. Sixty-one participants indicated that they were monolingual, fourteen indicated some knowledge of Spanish, seven indicated some knowledge of French, and others reported knowledge of a few other languages.

Materials. We constructed eighty-three nouns: forty-eight target items with nonlow vowels, and thirty-five fillers with [ɑ]. The forty-eight target items, listed in Appendix D, were made by crossing two nonlow vowels (e, u) * two shapes (trochee, iamb), and creating twelve items in each category. The unstressed vowel in each stem was schwa. The fillers consisted of seventeen trochees and eighteen iambs, with [ɑ] in their stressed syllable and schwa in their unstressed syllable.

The vowel inventory of the language, then, was [ɑ e i o u ə], which the participants saw spelled ⟨a e i o u –⟩; for example, [tɹəˈmel] was spelled ⟨tr–mel⟩.

The 'plurals' for the targets were created by flipping vowel backness; that is, singular [e] was switched with [o], and singular [u] was switched with [i], as shown in Table 5 below. For the fillers, the suffix [-ni] was added to make the plural, leaving the stem vowels unchanged, for example, [təˈkɑʃ ~ təˈkɑʃni], [ˈsɑgətʃ ~ ˈsɑgətʃni].

The items were recorded by a phonetically trained male native speaker of English in his twenties from Wisconsin, using the same procedure described in §5. To assign the experimental items plausible meanings, the same pictures were used as in §5.

Task. The experiment was delivered over the internet and presented as an artificial-grammar learning game. In the beginning of the experiment, the web server randomly assigned each participant to one of two training groups: trochaic or iambic.

The training stage consisted of ten trochaic target items with nonlow vowels (five [e], five [u]) in the trochaic training group, and ten such iambs for the iambic training group. Both groups were presented with six fillers with [ɑ], three iambs and three trochees. In the testing stage, all participants were presented with ten trochees (five [e], five [u]), ten iambs (five [e], five [u]), and ten fillers with [ɑ] (five iambs, five trochees).

---

the plural, it always impacts the initial stressed syllable, as in [ˈflus ~ ˈfly:sə] 'river' and [ˈbru:dər ~ ˈbry:dər] 'brother'. The diminutive [-çən] can impact noninitial syllables so long as they are stressed, for example, [admiˈral ~ admiˈrɛl-çən] 'admiral'.

	TROCHAIC TRAINING		IAMBIC TRAINING	
TRAINING	10 TROCHEES		10 IAMBS	
	= UMLAUT IMPACTS σ_1		= UMLAUT IMPACTS σ_2	
	ˈzuməp	ˈziməp	səˈfup	səˈfip
	ˈbɹezəl	ˈbɹozəl	tɹəˈmel	tɹəˈmol
	6 FILLERS		6 FILLERS	
	məˈfɑt	məˈfɑtni	məˈfɑt	məˈfɑtni
	ˈgɑstəʃ	ˈgɑstəʃni	ˈgɑstəʃ	ˈgɑstəʃni
TESTING	10 TROCHEES		10 TROCHEES	
	= UMLAUT IMPACTS σ_1		= UMLAUT IMPACTS σ_1	
	ˈfunəl	—	ˈfunəl	—
	ˈʃebəf	—	ˈʃebəf	—
	10 IAMBS		10 IAMBS	
	= UMLAUT IMPACTS σ_2		= UMLAUT IMPACTS σ_2	
	pəˈdul	—	pəˈdul	—
	kəˈzem	—	kəˈzem	—
	10 FILLERS		10 FILLERS	
	bɹəˈgɑm	—	bɹəˈgɑm	—
	ˈbɑləd	—	ˈbɑləd	—

TABLE 5. Artificial-grammar setup for experiment 4a.

The training and testing followed the same procedure as in §5, ending with fifteen trials where participants have to provide plurals for stems that they hear but do not see written, to ensure that they have listened to the audio materials.

RESULTS. To interpret the plurals that participants typed in, we removed any ⟨ni⟩ suffixes and all consonants, leaving only the vowel graphemes ⟨a e i u o⟩ and ⟨–⟩. Responses with one full vowel grapheme and one dash/schwa in the same order as in the singular were considered to match the singular's prosodic shape (69% of all responses), and thus potentially to be in compliance with the task. Plurals that had ⟨i⟩ in response to a singular [u] or ⟨o⟩ in response to a singular [e] were considered to have an umlauted vowel mapping, though their prosodic shape might not match the singular (32% of all responses). Finally, fully umlauted responses were those that matched the prosodic shape of the singular and also had an umlauted vowel mapping (28% of all responses). All other responses were considered nonumlauted.

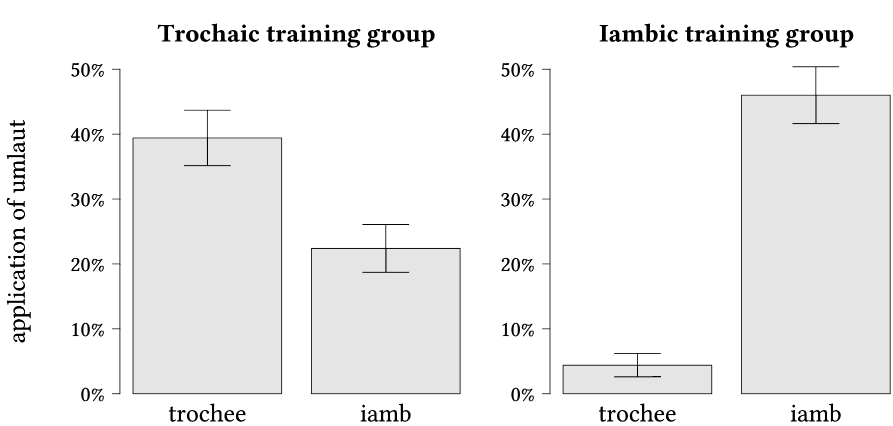


FIGURE 8. English speakers generalize cautiously to initial syllables, experiment 4a.

In the trochaic training group, umlauted (fully successful) responses were provided for trochaic items more than iambs (39% and 22%, respectively, a significant difference). In the iambic training group, umlauted responses were provided significantly more often for iambic items than for trochaic items (47% vs. 5%, respectively, a much larger significant difference), as seen in Figure 8. There is also a significant main effect of *group*, meaning that overall, umlaut was applied more in the group that received trochaic training.

The statistical analysis was performed with an *lmer* mixed-effects logistic regression model, again with the predictors *group* and *untrained* and their interaction, as in §5. The resulting model is reported in Table 6.[15]

	β	SE(β)	t	p-value
(intercept)	–2.10	0.22	–9.50	
group	–1.49	0.45	–3.30	< 0.001
untrained	–3.48	0.40	–8.62	< 0.0001
group:untrained	–3.67	0.79	–4.66	< 0.0001

TABLE 6. Artificial-grammar model for experiment 4a.

We see that participants in both groups apply the vowel alternation significantly less to the shape they were not trained on; that is, they extend the alternation cautiously. Of interest here is the significant interaction: participants were significantly more reluctant to extend the alternation in the iambic training group than they were in the trochaic training group. This is what we expected based on the theory developed here: with initial syllables being protected by default, participants are reluctant to impact the initial syllables of trochees if they have only seen the alternation impacting noninitial syllables in iambs.

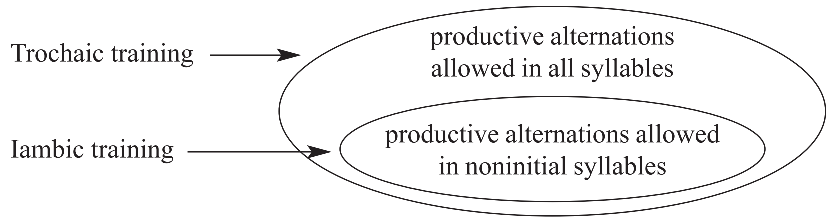
The two vowel mappings, [e → o] and [u → i], were applied equally well in both groups, at 28%. Naturally therefore, adding a vowel predictor to the model did not make a significant improvement.

6.3. DISCUSSION. In this experiment, participants learned the vowel changes [e → o] and [u → i] in stressed syllables. Unstressed syllables were kept as [ə] throughout. Covarying the stress and the position of the alternation allowed us to keep stress constant, with variation only in whether the alternation was in the initial or noninitial syllable.

Participants who were trained on iambs, where the alternation impacts the noninitial syllable, were strongly reluctant to impact initial syllables in the testing phase; this contrasts with those trained on trochees, where the alternation impacted the initial syllable, who were more willing to extend the alternation to the noninitial syllable. These results confirm our hypothesis that alternations in initial syllables license inferences to alternations in noninitial syllables, but not vice versa, as shown in the Venn diagram in 8. We conclude that the protection of monosyllables in §5 was a case of the more general protection of initial syllables.

[15] The model in Table 6 is a fully crossed model that had *group*, *untrained*, and their interaction as fixed effects, and the same predictors as random slopes for *item* and *participant*. This model uses all 2,000 data points (twenty nonlow vowel stimuli * 100 participants); the model that uses the 1,384 data points that match the prosodic shape of the singular is essentially identical. The model has low collinearity measures (VIF ≤ 3.25, $\kappa = 1.0002$).

(8)



We note a few differences between the results of experiments 3 and 4a. One difference has to do with the magnitude of the effects. In experiment 3, the increased reluctance to generalize to monosyllables in the iambic training group was significant, yet small in magnitude. In experiment 4a, the increased reluctance to generalize to initial syllables in the iambic training group is rather dramatic. Since the experiments differ in a number of respects, including the nature of the alternation involved, we hesitate to attribute this difference to any one factor.

A second quantitative (but not qualitative) difference has to do with the overall application of the alternation, which was the same in both groups in experiment 3, while experiment 4a showed a small yet significant overall bias toward applying the alternation in the trochaic training group. Referees have raised the concern that the effect is somehow related to vowel changes in the irregular pasts in English, which always impact the final syllable of the verb, as in *become* ~ *became*. This may bias English speakers to license vowel changes in iambs, and thus increase the overall application of umlaut. The suffixal nature of the alternation in this experiment may have played a role in eliciting past-tense-like responses as well, since the inflectional morphology of English is all suffixal, and thus more likely to impact final syllables. Following an editor's suggestion, the next experiment was designed to investigate the same alternation with a prefixal bias in order to explore this effect.

7. Experiment 4b: initial syllables also protected in polysyllables. In this artificial-grammar experiment, the same vocalic alternation of experiment 4a is shown based on a prefixal source. The iambic group in experiment 4a was trained on a local pattern (assuming that the plural was analyzed as a suffix, based on the fillers) and invited to extend it to a nonlocal case. The present experiment was designed to demonstrate that initial syllables were the conditioning factor between the two groups, and not locality, by basing the plural on a prefixal source. As we see below, in the current experiment, participants were attuned to a prefixal source for the plural, and thereby applied the alternation more to initial syllables, but initial syllables were still protected more strongly than noninitial syllables.

7.1. Materials and methods.

Participants. The participants ($N = 100$) were recruited through Amazon's Mechanical Turk and did not participate in any of the previous experiments, with fifty assigned to the trochaic training group and fifty assigned to the iambic training group. They were paid $1 for their time. The server logs indicate that these participants took on average eleven minutes to complete the experiment (range 6–37 minutes, median 10).

Participants had the option to volunteer demographic information. Year of birth was provided by all participants, who reported an average age of thirty-one (range 18–61, median 28). There was no significant age difference between the groups (difference in mean ~2 years, two-sample *t*-test, $t(97.35) = 0.71$, $p > 0.1$). Gender was reported by fifty-two females and forty-eight males. There was no significant gender difference between the groups (trochaic training: twenty-four females, iambic training: twenty-eight females, Fisher's exact test, $p > 0.1$). Ninety-eight participants said they grew up speaking English; two said they did not. As for the variety of English spoken, almost everybody named a state or a major city in the US or said they speak general American English. Forty-nine participants indicated that they were monolingual, eighteen indicated some knowledge of Spanish, fifteen indicated some knowledge of French, and others reported knowledge of a few other languages.

MATERIALS AND TASK. The materials were constructed exactly as in experiment 4a, but instead of the suffix [-ni], the plurals for the fillers have the PREFIX [ni-], for example, [təˈkɑʃ ~ nitəˈkɑʃ], [ˈsɑgətʃ ~ niˈsɑgətʃ]. The items were recorded by the same speaker from Pennsylvania who recorded the stimuli for experiment 3. The task was exactly as in experiment 4a.

7.2. RESULTS. The results were interpreted as in experiment 4a, except that ⟨ni⟩ was removed from the left edge of the responses rather than the right edge.

Responses with one full vowel grapheme and one dash/schwa in the same order as in the singular were considered to match the singular's prosodic shape (72% of all responses). Plurals that had ⟨i⟩ in response to a singular [u] or ⟨o⟩ in response to a singular [e] were considered to have an umlauted vowel mapping, though their prosodic shape might not have matched the singular (24% of all responses). Finally, fully umlauted responses were those that matched the prosodic shape of the singular and also had an umlauted vowel mapping (22% of all responses). All other responses were considered nonumlauted.

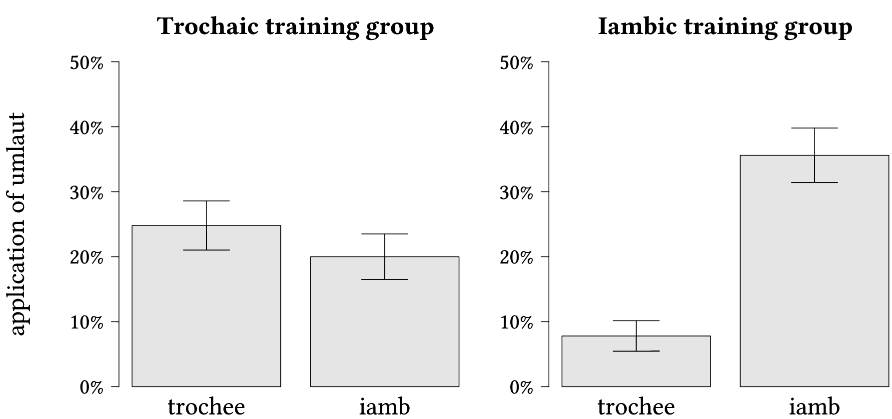


FIGURE 9. English speakers generalize cautiously to initial syllables, experiment 4b.

In the trochaic training group, umlauted (fully successful) responses were provided for trochaic items more than iambs (25% and 20%, respectively, a significant difference). In the iambic training group, umlauted responses were provided significantly more often for iambic items than trochaic items (36% vs. 8%, respectively, a larger significant difference), as seen in Figure 9. There were no significant differences between the groups in the overall application of umlaut.

The statistical analysis was performed as in experiment 4a. The resulting model is reported in Table 7.[16] We observe no significant difference in the overall rate of application of the alternation (no significant main effect for *group*), and in this respect the

[16] The model was built as in experiment 4a. It has low collinearity measures (VIF ≤ 1.52, κ = 1.0002).

results are different from those of experiment 4a. In other respects, the results are the same: participants in both groups apply the vowel alternation significantly less on the shape they were not trained on—that is, they extend the alternation cautiously, and the caution is significantly stronger in the iambic group. In other words, initial syllables are still protected more strongly than noninitial syllables, even when there is no overall preference for umlaut in either group.

	β	SE(β)	t	p-value
(intercept)	–2.52	0.21	–12.08	
group	–0.04	0.21	–0.21	> 0.1
untrained	–1.27	0.19	–6.64	< 0.0001
group:untrained	–0.46	0.19	–2.42	< 0.05

TABLE 7. Artificial-grammar model for experiment 4b.

7.3. DISCUSSION. In both experiment 4a and experiment 4b, participants applied umlaut to new prosodic types with caution, but the caution was stronger in the iambic training group. Both experiments therefore support the view that an alternation in a noninitial syllable does not license participants to extend the same alternation to initial syllables.

In experiment 4a, there was an overall preference for applying umlaut in the trochaic training group, which raised a concern about the possible influence of irregular past-tense verbs. In the present experiment, where the alternation was prefixal, the main effect disappeared, thereby successfully alleviating the concern about the main effect of *group* and squarely localizing the difference in the directionality of generalization to untrained items.

8. DISCUSSION: INITIAL-SYLLABLE FAITHFULNESS PROTECTS MONOSYLLABLES. As discussed in §2, most known cases of local morphophonological alternations impact elements in noninitial syllables more strongly than they impact the same stem-final elements in monosyllables, a result we obtained in four different experiments reported above. In this section, we interpret these results in light of the literature on initial-syllable faithfulness (Steriade 1994, Beckman 1997, 1998, Casali 1998, Barnes 2006, Becker 2009, Jesney 2009, Becker et al. 2011).

We provide an analysis in terms of optimality theory for concreteness, emphasizing the compatibility of this interpretation with a variety of other models. Given a set of constraints that includes both general faithfulness (IDENT) and a faithfulness constraint that only regulates against unfaithful mappings in the initial syllable (IDENT-σ_1), alongside a relevant markedness constraint (MARK), there are only three languages that are possible, as schematized in 9. With MARK outranking both faithfulness constraints, alternations will be observed in all stems, regardless of size. With general IDENT outranking MARK, no alternations will be observed, regardless of the ranking of IDENT-σ_1. When MARK is outranked by IDENT-σ_1 but not general IDENT, monosyllables are protected from alternations, but polysyllables are not. Crucially, this set of constraints cannot describe a pattern in which monosyllables are more strongly impacted by markedness than polysyllables.

(9) MARK ≫ IDENT, IDENT-σ_1: Alternations observed everywhere
IDENT, IDENT-σ_1 ≫ MARK: Alternations observed nowhere
IDENT-σ_1 ≫ MARK ≫ IDENT: Alternations observed in polysyllables

The results of the ranking possibilities in 9 are important, because they demonstrate that once the learner brings to the task an inventory of constraints with specialized reference to initial syllables but without any specialized reference to noninitial syllables, it will not be possible to encode (and therefore generalize) a pattern of greater alternation in initial syllables as opposed to noninitial syllables. This is precisely what we observed with the English fricative voicing pattern: it cannot be described in terms of 9, and was not extended either in wug tasks or artificial-grammar tasks.

To complete our discussion, we outline an implementation in which the predictions made in 9—stated in categorical, exceptionless terms—are not disrupted by the presence of the lexical exceptions. We demonstrate that the basic pattern holds when supplemented with the USELISTED approach to exceptionality (Zuraw 2000, Hayes & Londe 2006; see also Zhang & Lai 2008, Zhang et al. 2011), noting that many other approaches to exceptionality may be compatible as well.

USELISTED is a constraint that regulates against the productive combination of stems and affixes when learners already have the combination listed in their lexicon. For example, for an adult English speaker, a common plural like *knives* is protected by USELISTED, as shown in 10. Candidates (a) and (b) were generated directly from the listed form [naɪvz], whereas candidates (c) and (d) are made by combining the root [naɪf] with the plural suffix. The constraints IDENT and IDENT-σ_1, which here only regulates against unfaithful mappings in the root, prefer the faithful fricative (voiceless in the underlying representation, voiced in the listed plural). Markedness (specifically, the lenition or voicing assimilation in the plural) prefers the voiced fricative, but both faithfulness and markedness are outranked by USELISTED.

(10) *knives* protected by USELISTED

/naɪf + z/, listed: [naɪvz]	USELISTED	IDENT	IDENT-σ_1	MARK
a. ☞ naɪvz → naɪvz				
b. naɪvz → naɪfs		*	*	*
c. naɪf + z → naɪvz	*!	*	*	
d. naɪf + z → naɪfs	*!			*

With a nonce word such as [bɹaɪf], the speaker has no plural form listed, and combining the base and the plural suffix via the grammar (without listed exceptionality) is the only option, as shown in 11. Both candidates violate USELISTED equally, and the decision is thereby handed down to markedness and faithfulness constraints, which may be stochastically ranked or weighted for analyzing nonce words.

(11) *brife* controlled by markedness and faithfulness

/bɹaɪf + z/, listed: []	USELISTED	IDENT	IDENT-σ_1	MARK
a. ☞ bɹaɪf + z → bɹaɪvz	*	*	*	
b. ☞ bɹaɪf + z → bɹaɪfs	*			*

USELISTED will be similarly silent for a novel polysyllable such as [səˈbaɪf], which differs from [bɹaɪf] in that a voicing alternation only incurs a violation of IDENT, not IDENT-σ_1. This means that regardless of the ranking/weighting of IDENT-σ_1, [səˈbaɪvz] has to be at least as acceptable as [bɹaɪvz]. There is thus no way for the grammar to encode a tendency to voice more in monosyllables.

(12) *sebife* controlled by markedness and general faithfulness

/səˈbaɪf + z/, listed: []	USELISTED	IDENT	IDENT-σ_1	MARK
a. ☞ səˈbaɪf + z → səˈbaɪvz	*	*		
b. ☞ səˈbaɪf + z → səˈbaɪfz	*			*

An analysis employing listed exceptions can enforce faithfulness to any arbitrary list of plurals in the memorized lexicon, while leaving the productivity of such alternations for novel forms completely up to the grammar, where constraints specific to initial syllables may step in.[17]

Importantly, however, for nonce words (as in a wug task, or any generalization to novel data), the decision is arbitrated purely grammatically, between markedness and faithfulness, and therefore the range of possible generalizations is limited to the one schematized in 9. This effect corresponds to what was observed with English speakers in experiments 1 and 2: the lexicon contains a counteruniversal arrangement of existing nouns, with more alternations in monosyllables, but the nonce words exhibit the opposite pattern, in concord with the possibilities in 9.

We leave the exact nature of the markedness constraint that causes the alternation in English somewhat open, as we have studied the morpholexical distribution of the alternation and not its phonetic realization. While English contrasts two kinds of fricatives word-finally, as in [tɹuθ] 'truth' vs. [smuð] 'smooth', the transcription does not do justice to the realization of the contrast, which involves a short vowel and long fricative in *truth*, but a longer vowel and shorter, weaker, optionally voiceless fricative in *smooth*. Since it is not clear that the process generates a genuinely voiced fricative, it is not clear that this relates to the debate about final voicing (Blevins 2004, Yu 2004, Kiparsky 2006, et seq.). We conjecture that the 'voicing' alternation is a case of morphologically restricted lenition, which historically was more widespread in English (see Honeybone & Spaargaren 2011 and references within).[18] In fact, connecting the alternation to the differences in duration heard both on the fricative itself and on the preceding vowel makes sense of the effect of vowel length: leniting a fricative following a long vowel like [i] further lengthens that vowel and enhances its length. By contrast, leniting a fricative following a short vowel like [ɪ] lengthens the vowel, thus obscuring its quantitative opposition as a short vowel.

In the constraint-based approach outlined here, the propensity for voicing after long vowels is modeled as a markedness effect: for example, one or more constraints against voiceless fricatives following a long vowel, limited to the appropriate morphological context. Voicing is blocked in unstressed syllables by a constraint against voiced codas in unstressed syllables; this constraint has few exceptions in English in general. This markedness constraint has the side effect of causing more faithful outcomes in the unstressed position, and thus appears to conflict with constraints that require faithfulness to the stressed position, but note that such conflicts are unremarkable in theories based on the interaction of violable constraints.

To summarize, then, we have seen in this section how the initial-syllable faithfulness approach accounts for the patterns of experiments 2–4. The factorial typology generates

[17] While this result fundamentally depends on the restricted ranking possibilities offered by the set of constraints in 9, the treatment of exceptions alongside that grammar need not be limited to the UseListed approach. The analysis of lexical exceptions in terms of constraint cloning (Pater 2006, 2009, Coetzee 2008, Becker 2009, Becker et al. 2011) would provide the same result: existing items are listed with clones of the constraints in 9, with nonce forms again giving rise to one of the universally available generalizations.

[18] We have treated the laryngeal alternation as a lenition process that impacts fricatives. Kie Zuraw (p.c.) suggests exploring whether the markedness constraint involves voicing assimilation applying to fricatives, triggered by the plural suffix. Given the arguments that the English laryngeal opposition is based on [spread glottis] rather than [voice] (Kager et al. 2007), as well as the more general patterning of the lenition alternation in derived forms such as *bath ~ bathe* and *thief ~ thievery*, we uphold the analysis in the text, noting however that future work might arbitrate between these analyses by looking into the relation between alternations in the plural and in derived nouns, verbs, and adjectives.

languages with alternations that apply to monosyllables and polysyllables equally, or languages with alternations that apply to polysyllables only, but no languages with alternations applying to monosyllables only. While lexical exceptions can be found in a language like English, where the alternation applies to the existing listed monosyllables more than it does to existing listed polysyllables, this difference does not make it into a grammatical encoding that can be generalized.

9. CONCLUSIONS. Experiment 1 found that in the English lexicon, monosyllables are impacted by voicing alternations (as in *knife ~ knives*) more than polysyllables. In fact, the top third of the scale in our rating task was populated by monosyllables only. Iambs and trochees (defined here as polysyllables with a final stressed and unstressed syllable, respectively) were less acceptable with voicing alternations. In experiment 2, a nonce word task (wug test), however, monosyllables and iambs were equally acceptable as voiced by English speakers. This means that a pattern that is true of the lexicon was not generalized to novel words, illustrating a 'surfeit of the stimulus' effect (also known as 'Orwell's problem' in Chomsky 1986, though in a different context). Such results underscore the necessity of complementing typological work with experimental work, since some languages may offer generalizations in the lexicon that do not make it into the speakers' grammar. Omnivorous statistical learners must be restrained by a set of built-in biases, or they will fail to model the grammar that learners arrive at. As Zimmer (1969:309) aptly puts it:

> It is occasionally assumed that, if a regularity can be stated, this alone permits us to infer some kind of psychological reality; but there is surely nothing necessary about this assumption. One might equally well assume that someone who learns the sequence of numbers 1, 5, 19, 65, 211, 665 must necessarily know the formula which relates them (namely that the n^{th} member of the series equals $3^n - 2^n$). Of course, if a person could not only repeat the sequence correctly, but also continued it on his own with 2059, that would be evidence that he knows the formula in question; but it is just such conclusive evidence that is lacking in the case of certain regularities found in languages.

Although English learners may have memorized the patterns of fricative voicing in monosyllables, we connect their inability to extend this pattern to novel words to a universal implicational relationship: alternations in monosyllables imply alternations in polysyllables, but alternations in polysyllables allow monosyllables to remain protected, as observed in Turkish and in other languages. By hypothesis, the grammar cannot encode a preference for more alternations in monosyllables than polysyllables, and if there is such a pattern in the lexicon of a language, it cannot be learned. We derive this implicational relationship from initial-syllable faithfulness, which states that the word-initial syllable is protected from alternations. Constraints that protect NONinitial syllables are absent from CON, thus restricting the range of possible grammars. We note that while much recent literature explores the idea of inducing markedness constraints from the ambient language (e.g. Boersma & Pater 2007, Hayes & Wilson 2008, Moreton 2010), we know of no such proposals for learning faithfulness constraints. It would seem at the moment that faithfulness constraints, including their sensitivity to strong positions (e.g. onset, stressed syllable, initial syllable), are still a part of what the learner brings to the table.

To show that even English speakers are biased to prefer the protection of monosyllables, we conducted three artificial-grammar experiments. In the first, we trained half of our participants on alternations in monosyllables, and half on alternations in polysyllables. While both groups applied the alternation cautiously to the shape they had not seen, those trained on polysyllables applied the alternation significantly more cautiously to the monosyllables they had not seen, relative to those trained on monosyllables. In other words, the English speakers deployed the universal implicational relationship seen in Turkish.

To strengthen the connection between the monosyllabicity effect and initial-syllable faithfulness, we conducted two artificial-grammar experiments in which all stems were disyllabic. We trained half of our participants on alternations in trochees, and half on iambs. Both groups of participants were cautious in applying the alternation to the position they had not seen it in, but those trained in applying the alternation in iambs were significantly more cautious than those trained on trochees. The same kind of bias ran through all three experiments: initial syllables are protected, and learning alternations in noninitial syllables does not license the inference to apply the alternations to initial syllables.

The set of results that English speakers do not extend the countertypological pattern from their lexicon, and prefer the Turkish-like pattern in an artificial language, is inconsistent with Barnes's (2006) suggestion that initial-syllable faithfulness is wholly due to increased duration of initial syllables. As Barnes (2006) notes, initial syllables have longer duration in Turkish and shorter duration in English, and yet we find that speakers of both languages prefer the protection of initial syllables. It seems doubtful, then, that the protection of monosyllables is purely phonetically grounded. Deriving the protection of monosyllables from properties of lexical neighborhoods, as proposed in Ussishkin & Wedel 2009, has the potential advantage of predicting a uniformly crosslinguistic preference for less alternations in short words, since neighborhood density is always inversely correlated with length. However, the neighborhood density effect is shown to be too underpredictive in Pycha et al. 2007, Becker & Nevins 2009, and Becker et al. 2012. Furthermore, neighborhood density will not be able to distinguish local and nonlocal alternations, potentially leaving Chamorro umlaut unaccounted for. We conclude that at present, the protection of monosyllables by initial-syllable faithfulness must be an atomic, symbolic element of the theory, though we leave open the possibility that the functional motivation has been phonologized (Hyman 1976, 2008) and incorporated into the grammatical apparatus.

We have now gathered diverse sources of evidence that initial-syllable faithfulness, and perhaps positional faithfulness more generally, is a powerful organizing force in the regulation of morphophonological alternations on stems. Since we see its effect even in learners who possess evidence to the contrary in their own language, we must characterize it as an inherent analytic bias, and thereby part of what constitutes universal grammar. We have shown that this bias causes a lexical generalization to go unlearned (a SURFEIT of the stimulus), and that it shapes the expectations that speakers have in artificial-grammar experiments with partial training data (a POVERTY of the stimulus). Unlike those positional faithfulness constraints that may be phonetically based (faithfulness to onsets, or stressed syllables), or positional faithfulness constraints that are sensitive to the morphology (faithfulness to roots, or nouns; Smith 2002, 2010), initial-syllable faithfulness seems to be a purely formal and purely phonological element of the theory. While it may be grounded in phonetic and psycholinguistic pressures, its status is phonologized, and it exerts a formal guiding hand in hypothesis formation during acquisition, as illustrated in the subset principle schema: it asymmetrically biases learners against generalizing from noninitial to initial syllables, but not vice versa.

Appendix A: Experiment 1 results by item

The items are listed with average responses (1 = voiceless, 7 = voiced) from 200 participants.

MONOSYLLABLES				IAMBS		TROCHEES	
bath	4.6	nymph	2.1	aftermath	3.5	absinth	3.0
berth	3.4	oaf	2.0	behalf	4.6	azimuth	2.5
bluff	1.6	oath	4.6	belief	1.7	bailiff	2.7
booth	3.5	path	5.4	blacksmith	2.8	Behemoth	4.0
breath	1.6	prof	1.6	carafe	2.9	billionth	3.2
brief	1.2	proof	2.4	castoff	2.6	caliph	2.3
broth	3.5	puff	2.0	coelacanth	3.5	eightieth	3.2
calf	6.9	quaff	2.4	earmuff	1.9	goliath	2.1
chef	1.9	reef	2.8	eighteenth	2.3	hieroglyph	2.0
chief	3.4	ref	1.1	epitaph	3.2	hyacinth	3.0
clef	2.1	roof	3.4	giraffe	3.6	kerchief	3.9
cliff	2.0	safe	2.4	handcuff	2.5	mammoth	3.2
cloth	4.0	scarf	6.9	midriff	2.0	mastiff	2.8
cough	1.7	self	6.7	motif	2.9	monolith	2.7
cuff	1.4	serf	1.3	paragraph	2.5	plaintiff	2.9
death	1.7	sheaf	5.8	pilaf	2.9	pontiff	2.5
dwarf	6.3	sheath	5.2	polymath	3.4	Sabbath	3.7
earth	3.3	shelf	6.9	psychopath	4.0	seraph	3.2
elf	7.0	skiff	2.3	relief	3.9	serif	2.6
faith	2.8	sleuth	3.9	sabertooth	3.2	sheriff	2.1
fife	2.4	sloth	2.7	vermouth	4.7	tariff	2.2
fourth	3.2	sniff	1.5			triumph	2.2
gaffe	2.1	spoof	2.6			zenith	3.6
goof	2.0	staff	2.1				
Goth	3.6	stiff	1.0				
graph	1.6	strength	1.4				
growth	1.8	swath	4.6				
gulf	2.8	thief	6.6				
half	6.8	tiff	2.7				
hearth	3.2	trough	3.4				
heath	2.4	truth	4.2				
hoof	6.8	turf	2.5				
knife	6.0	waif	2.5				
laugh	2.3	wharf	5.1				
leaf	6.6	whiff	2.5				
life	7.0	width	3.7				
loaf	6.3	wife	6.4				
month	1.7	wolf	6.6				
moth	4.8	wraith	4.7				
mouth	4.6	wreath	4.6				
myth	2.2	youth	4.0				

Appendix B: Experiment 2 results by item

The items are listed with average responses (1 = voiceless, 7 = voiced) from 200 participants.

MONOSYLLABLES		IAMBS		TROCHEES	
beɪf	5.7	bəwælf	4.9	bɹʌnləf	3.9
bɹaɪf	5.5	glənɑf	4.6	ʤɑnəf	3.9
bʌnf	4.5	həbɑf	4.8	floʊdəf	4.5
dif	6.0	jəstɚf	5.1	glɚsəf	4.3
duf	3.5	kəzɪnf	3.3	junəf	4.3
ʤæf	3.8	kwəzʌf	4.1	jælməf	3.6

(*continues*)

MONOSYLLABLES		IAMBS		TROCHEES	
glif	5.2	məleɪf	5.5	kaɹəf	3.4
jɛf	4.2	nədʌlf	4.6	kɹeɪdəf	5.0
kɚf	3.5	nəʃɪf	5.8	nɛlgəf	4.3
klʊf	5.1	nətoʊf	5.2	pidəf	4.9
meɪf	5.6	pətɛlf	5.6	ɹɪlkəf	3.8
nɪf	3.7	pɚnæf	5.1	ɹaɪləf	4.3
noʊf	4.9	ɹədif	5.2	seɪləf	4.2
ɹoʊf	5.2	səbaɪf	5.3	stɚnəf	4.5
sɛnf	4.4	səkloʊf	5.8	stʊkəf	4.5
smɑf	5.0	stəɹuf	5.1	ʃædəf	3.8
stɚf	5.0	ʃənɚf	4.5	ʃɛstəf	5.7
sʌf	3.4	təŋkeɪf	5.2	ʃʌtəf	3.6
ʃɪlf	5.2	tɹəluf	5.1	tɑkəf	4.6
wɑf	3.0	wəgɛf	4.3	ʧɪdəf	3.9
zælf	5.1	wəɹif	5.4	voʊləf	4.4
zaɪf	5.1	zəkaɪf	4.8	wibəf	4.0
biθ	5.0	bɹəloʊθ	5.1	ʤʌkəθ	4.2
bɪlθ	4.2	dəkɛlθ	4.6	fɪdəθ	2.6
doʊθ	5.5	dəpaɪθ	4.4	haɪləθ	3.8
dwiθ	5.3	ʤəzɑθ	3.4	hɑkəθ	2.9
huθ	5.3	gənɚθ	4.1	kælnəθ	3.2
kɹæθ	4.1	həleɪθ	5.1	lɚbəθ	3.6
kʌθ	3.9	həɹiθ	4.5	lisəθ	3.9
kwɚθ	4.0	həsaɪθ	4.2	moʊdəθ	3.2
lɪθ	3.7	kənɑθ	3.8	nækəθ	3.5
pælθ	4.9	kənuθ	4.6	neɪgəθ	3.3
pɑɹθ	4.0	nəboʊθ	4.5	paɪdəθ	2.5
peɪθ	4.4	ɹəstɪlθ	5.3	ɹoʊsəθ	4.3
pɛlθ	3.8	ɹəstɛθ	4.3	ɹuvəθ	3.0
pɹʊθ	4.1	səkɚθ	4.3	ɹikɹəθ	3.9
skɛθ	4.5	səpʌlθ	4.8	sɑstəθ	4.2
smeɪθ	4.8	səsælθ	5.2	sɛŋgɪθ	3.4
stɚθ	3.2	səstʌθ	3.6	sɪltəθ	3.6
sʌlθ	3.6	ʃədʊθ	3.9	speɪləθ	3.4
ʃɹaɪθ	4.9	ʃəkæθ	5.0	ʃɛnəθ	3.6
ʧaɪθ	5.2	təɹeɪθ	4.0	tʌlpəθ	3.6
ʧoʊθ	4.9	təspɪθ	3.5	wɚgəθ	2.3
wɑθ	4.2	zəziθ	3.9	wʊkəθ	3.6

Appendix C: Experiment 3 results by item

The items are listed with the average responses in each group (0 = voiceless, 1 = voiced) from 100 participants (fifty in each group).

	MONO TRAINING	IAMB TRAINING		MONO TRAINING	IAMB TRAINING
zɑp	.60	.46	dəɹɑp	.60	.57
flep	.72	.42	həsɑp	.76	.67
sep	.47	.44	təgep	.52	.65
mip	.67	.44	kənip	.42	.50
bop	.74	.43	ɹəlip	.47	.62
ʃop	.58	.63	fəʧop	.64	.67
nup	.67	.36	gətop	.65	.59
ʧup	.70	.76	ʤəlup	.65	.71
bɹaɪp	.64	.42	səɹup	.61	.62
gaɪp	.80	.56	ʧənaɪp	.69	.53

(*continues*)

	MONO TRAINING	IAMB TRAINING
gɹɑt	.65	.57
dɹet	.71	.47
klet	.36	.43
vit	.65	.45
zot	.62	.31
ʒot	.75	.55
pɹut	.50	.38
stut	.53	.53
paɪt	.82	.42
skaɪt	.46	.21
stɹɑk	.64	.47
glek	.65	.46
ʧek	.44	.47
nik	.46	.47
dok	.45	.25
pɹok	.53	.41
fuk	.48	.40
huk	.65	.27
blaɪk	.55	.46
ʃaɪk	.45	.38

	MONO TRAINING	IAMB TRAINING
fɹədɑt	.59	.55
nəmɑt	.67	.60
ʒənet	.50	.47
bəgit	.35	.60
səplit	.56	.31
ʃəlot	.50	.55
ʧədot	.64	.50
gəʃut	.78	.44
kləput	.48	.47
fəhaɪt	.50	.50
gləzɑk	.33	.46
ʧəpɑk	.31	.55
pənek	.92	.57
bəsik	.44	.56
gəfɹok	.50	.62
ləʃok	.44	.67
bəʧuk	.57	.23
ʒətuk	.47	.50
ɹədaɪk	.73	.53
səpaɪk	.69	.62

Appendix D: Experiment 4a results by item

The items are listed with the average responses in each group (0 = nonalternating vowel, 1 = umlauted) from 100 participants (fifty in each group).

	TROCHEE TRAINING	IAMB TRAINING
bəʧuk	.17	.21
ʧəmum	.20	.52
dɹətup	.39	.67
fɹəkuf	.23	.28
gəsut	.32	.47
kənul	.33	.45
kləput	.32	.57
pədul	.50	.48
pləsum	.17	.40
səfup	.11	.67
ʃɹətuk	.32	.62
təʃuf	.35	.50
bəndef	.38	.32
kəʃel	.13	.33
kəzem	.65	.36
məpek	.16	.53
nədep	.23	.36
pəsnem	.18	.47
pləzek	.29	.50
skənet	.14	.68
sməɹef	.25	.56
stəfet	.07	.62
təgep	.16	.73
tɹəmel	.23	.44

	TROCHEE TRAINING	IAMB TRAINING
blukəm	.20	.06
bulək	.38	.05
funəl	.46	.15
gunək	.46	.00
kuʃət	.35	.00
lupət	.38	.08
mustəp	.53	.11
pɹusəm	.46	.05
ʃundəf	.56	.14
suməl	.53	.12
sutɹəf	.47	.10
zuməp	.52	.06
bɹezəl	.44	.20
ʧedɹəl	.25	.00
deʃəf	.44	.10
fɹenəp	.42	.08
kefət	.39	.27
kɹefəm	.59	.04
kɹeʃək	.48	.05
lefəp	.42	.00
pɹekət	.35	.17
sefək	.32	.15
ʃebəf	.43	.05
zedɹəm	.43	.00

Appendix E: Experiment 4b results by item

The items are listed with the average responses in each group (0 = nonalternating vowel, 1 = umlauted) from 100 participants (fifty in each group).

	TROCHEE TRAINING	IAMB TRAINING		TROCHEE TRAINING	IAMB TRAINING
bətʃuk	.13	.50	blukəm	.28	.04
tʃəmum	.15	.22	bulək	.20	.04
dɹətup	.16	.36	funəl	.22	.05
fɹəkuf	.18	.57	gunək	.22	.09
gəsut	.24	.70	kuʃət	.33	.14
kənul	.15	.07	lupət	.24	.09
kləput	.20	.35	mustəp	.11	.00
pədul	.29	.19	pɹusəm	.29	.06
pləsum	.25	.57	ʃundəf	.38	.10
səfup	.26	.48	suməl	.38	.10
ʃɹətuk	.20	.39	sutɹəf	.17	.06
təʃuf	.22	.30	zuməp	.35	.24
bəndef	.40	.06	bɹezəl	.27	.05
kəʃel	.28	.29	tʃedɹəl	.10	.05
kəzem	.06	.50	deʃəf	.14	.05
məpek	.23	.43	fɹenəp	.41	.00
nədep	.22	.36	kefət	.24	.00
pəsnem	.14	.27	kɹefəm	.25	.20
pləzek	.17	.35	kɹeʃək	.22	.06
skənet	.22	.12	lefəp	.20	.12
sməɹef	.28	.13	pɹekət	.26	.10
stəfet	.09	.41	sefək	.22	.12
təgep	.21	.33	ʃebəf	.18	.08
tɹəmel	.18	.33	zedɹəm	.22	.05

REFERENCES

Alber, Birgit. 2001. Maximizing first positions. *Linguistics in Potsdam (Proceedings of HILP 5)* 12.1–19.

Albright, Adam, and Bruce Hayes. 2002. Modeling English past tense intuitions with minimal generalization. *Proceedings of the sixth meeting of the ACL Special Interest Group in Computational Phonology*, ed. by Michael Maxwell, 58–69. Philadelphia: Association for Computational Linguistics.

Barnes, Jonathan. 2006. *Strength and weakness at the interface: Positional neutralization in phonetics and phonology*. Berlin: Mouton de Gruyter.

Bates, Douglas, and Martin Maechler. 2009. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32. Online: http://cran.r-project.org/web/packages/lme4/index.html.

Becker, Michael. 2009. *Phonological trends in the lexicon: The role of constraints*. Amherst: University of Massachusetts Amherst dissertation.

Becker, Michael; Lauren Eby Clemens; and Andrew Nevins. 2012. A richer model is not always more accurate: The case of French and Portuguese plurals. Amherst: University of Massachusetts, ms. Online: http://ling.auf.net/lingBuzz/001336.

Becker, Michael; Nihan Ketrez; and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87.84–125.

Becker, Michael, and Andrew Nevins. 2009. Initial-syllable faithfulness as the best model of word-size effects in alternations. Paper presented at the 40th conference of the North East Linguistic Society (NELS), MIT, November 2009.

Beckman, Jill. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14.1–46.

BECKMAN, JILL. 1998. *Positional faithfulness*. Amherst: University of Massachusetts Amherst dissertation.

BERKO, JEAN. 1958. The child's learning of English morphology. *Word* 14.150–77.

BERWICK, ROBERT C. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.

BLEVINS, JULIETTE. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

BLUST, ROBERT. 2004. Austronesian nasal substitution: A survey. *Oceanic Linguistics* 43.73–148.

BOERSMA, PAUL, and JOE PATER. 2007. Constructing constraints from language data: The case of Canadian English diphthongs. Paper presented at the 38th conference of the North East Linguistic Society (NELS), University of Ottawa, October 2007.

BOERSMA, PAUL, and DAVID WEENINK. 2011. Praat: Doing phonetics by computer. Online: http://www.praat.org.

BYBEE, JOAN. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10.425–55.

CASALI, RODERIC. 1998. *Resolving hiatus*. New York: Garland.

CHOMSKY, NOAM. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

CHRISTDAS, PRATHIMA. 1988. *The phonology and morphology of Tamil*. Ithaca, NY: Cornell University dissertation.

CHUNG, SANDRA. 1983. Transderivational relationships in Chamorro phonology. *Language* 59.35–66.

COETZEE, ANDRIES W. 2008. Grammaticality and ungrammaticality in phonology. *Language* 84.218–57.

DELILKAN, ANN. 2005. Head-dependent asymmetry: Feet and fusion in Malay. *SKY Journal of Linguistics* 15.47–90.

ERNESTUS, MIRIAM, and R. HARALD BAAYEN. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79.5–38.

FANSELOW, GISBERT, and CAROLINE FÉRY. 2002. Ineffability in grammar. *Resolving conflicts in grammars: Optimality theory in syntax, morphology, and phonology* (Special issue of *Linguistische Berichte* 11), ed. by Gisbert Fanselow and Caroline Féry, 265–304. Tübingen: Helmut Buske.

GIAVAZZI, MARIA. 2010. *The phonetics of metrical prominence and its consequences for segmental phonology*. Cambridge, MA: MIT dissertation.

GOMES, CHRISTINA ABREU, and CAROLINA GONÇALVES MANOEL. 2010. Flexão de número na gramática de criança e na gramática do adulto. *Veredas On Line* 304056.122–34. Online: http://www.ufjf.br/revistaveredas/files/2010/08/ARTIGO-91.pdf

GOUSKOVA, MARIA, and MICHAEL BECKER. 2012. Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory*, to appear.

GRAVES, MICHAEL F., and STEPHEN KOZIOL. 1971. Noun plural development in primary grade children. *Child Development* 42.116–73.

HAY, JENNIFER; JANET PIERREHUMBERT; and MARY BECKMAN. 2004. Speech perception, well-formedness and the statistics of the lexicon. *Phonetic interpretation: Papers in laboratory phonology 6*, ed. by John Local, Richard Ogden, and Rosalind Temple, 58–74. Cambridge: Cambridge University Press.

HAYES, BRUCE. 1995. On what to teach the undergraduates: Some changing orthodoxies in phonological theory. *Linguistics in the morning calm 3*, ed. by Ik-Hwan Lee, 59–77. Seoul: Hanshin.

HAYES, BRUCE. 2004. Phonological acquisition in optimality theory: The early stages. *Constraints in phonological acquisition*, ed. by René Kager, Joe Pater, and Wim Zonneveld, 157–203. Cambridge: Cambridge University Press.

HAYES, BRUCE. 2009. *Introductory phonology*. Oxford: Wiley-Blackwell.

HAYES, BRUCE, and ZSUZSA LONDE. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23.59–104.

HAYES, BRUCE, and COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.

HAYES, BRUCE; KIE ZURAW; PÉTER SIPTÁR; and ZSUZSA LONDE. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85.822–63.

HONEYBONE, PATRICK, and MARLEEN SPAARGAREN. 2011. Not another case of final obstruent voicing? Laryngeal realism, Late Middle English and impossible phonological change. Handout for talk given at the Old World Conference on Phonology (OCP) 8, Marrakesh, January 2011.

HUALDE, JOSÉ IGNACIO. 1989. Autosegmental and metrical spreading in the vowel-harmony systems of Northwestern Spain. *Linguistics* 27.773–805.

HUBACK, ANA PAULA DA SILVA. 2007. *Efeitos de freqüência nas representações mentais*. Belo Horizonte: Universidade Federal de Minas Gerais dissertation.

HYMAN, LARRY. 1976. Phonologization. *Linguistic studies presented to Joseph H. Greenberg*, ed. by Alphonse Juilland, 407–18. Saratoga, CA: Anna Libri.

HYMAN, LARRY. 2008. Enlarging the scope of phonologization. *UC Berkeley Phonology Lab Annual Report* 2008.382–409.

INKELAS, SHARON, and CEMIL ORHAN ORGUN. 1995. Level ordering and economy in the lexical phonology of Turkish. *Language* 71.763–93.

INKELAS, SHARON; CEMIL ORHAN ORGUN; and CHERYL ZOLL. 1997. The implications of lexical exceptions for the nature of the grammar. *Derivations and constraints in phonology*, ed. by Iggy Roca, 393–418. Oxford: Clarendon.

JESNEY, KAREN. 2009. Positional faithfulness, non-locality, and the harmonic serialism solution. *North East Linguistic Society (NELS)* 39, to appear.

JESPERSEN, OTTO. 1909. *A modern English grammar on historical principles, part 1: Sounds and spellings*. Heidelberg: Carl Winter.

KAGER, RENÉ, and JOE PATER. 2012. Phonotactics as phonology: Knowledge of a complex restriction in Dutch. *Phonology* 29.1.81–111.

KAGER, RENÉ; SUZANNE VAN DER FEEST; PAULA FIKKERT; ANNEMARIE KERKHOFF; and TANIA S. ZAMUNER. 2007. Representations of [voice]: Evidence from acquisition. *Voicing in Dutch*, ed. by Jeroen van de Weijer and Erik Jan van der Torre, 41–80. Amsterdam: John Benjamins.

KALLESTINOVA, ELENA. 2004. Voice and aspiration of stops in Turkish. *Folia Linguistica* 38.117–43. (Special issue on voice, ed. by Grzegorz Dogil.)

KIPARSKY, PAUL. 1982. Word-formation and the lexicon. *Proceedings of the 1982 Mid-America Linguistics Conference*, ed. by Francis Ingemann, 3–29. Lawrence: University of Kansas.

KIPARSKY, PAUL. 2006. The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32.217–36.

LEWIS, GEOFFREY L. 1967. *Turkish grammar*. Oxford: Clarendon.

LISKER, LEIGH. 1974. On 'explaining' vowel duration variation. *Glossa* 8.233–46.

MANZINI, RITA, and KENNETH WEXLER. 1987. Parameters, binding theory, and learnability. *Linguistic Inquiry* 18.413–44.

MODER, CAROL LYNN. 1992. *Productivity and categorization in morphological classes*. Buffalo: SUNY Buffalo dissertation.

MORETON, ELLIOTT. 2008. Analytic bias and phonological typology. *Phonology* 25.83–127.

MORETON, ELLIOTT. 2010. Connecting paradigmatic and syntagmatic simplicity bias in phonotactic learning. Talk given at MIT, April 9.

MORETON, ELLIOTT, and JOE PATER. 2011. Learning artificial phonology: A review. Chapel Hill: University of North Carolina, and Amherst: University of Massachusetts, MS. Online: http://people.umass.edu/pater/moreton-pater-review.pdf.

PATER, JOE. 2006. The locus of exceptionality: Morpheme-specific phonology as constraint indexation. *Papers in optimality theory 3* (University of Massachusetts occasional papers in linguistics 32), ed. by Leah Bateman and Adam Werle, 1–36. Amherst: GLSA Publications.

PATER, JOE. 2009. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. *Phonological argumentation: Essays on evidence and motivation*, ed. by Steve Parker, 1–33. London: Equinox.

PETROVA, OLGA; ROSEMARY PLAPP; CATHERINE RINGEN; and SZILÁRD SZENTGYÖRGYI. 2006. Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review* 23.1–35.

PRASADA, SANDEEP, and STEVEN PINKER. 1993. Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes* 8.1–56.

PYCHA, ANNE; SHARON INKELAS; and RONALD SPROUSE. 2007. Morphophonemics and the lexicon: A case study from Turkish. *Experimental approaches to phonology*, ed. by Maria-Josep Solé, Patrice Speeter Beddor, and Manjari Ohala, 369–85. Oxford: Oxford University Press.

R DEVELOPMENT CORE TEAM. 2011. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org.

SCHNOEBELEN, TYLER, and VICTOR KUPERMAN. 2010. Using Amazon Mechanical Turk for linguistic research. *Psihologija* 43.419–28.

SELKIRK, ELISABETH O. 1982. *The syntax of words*. Cambridge, MA: MIT Press.

SMITH, JENNIFER L. 2000. Positional faithfulness and learnability in optimality theory. *Eastern States Conference on Linguistics (ESCOL '99)* 14.203–14.

SMITH, JENNIFER L. 2002. *Phonological augmentation in prominent positions*. Amherst: University of Massachusetts Amherst dissertation.

SMITH, JENNIFER L. 2010. Category-specific effects. *The Blackwell companion to phonology*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 2439–63. Oxford: Wiley-Blackwell.

SPROUSE, JON. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43.155–67.

STAUSLAND JOHNSEN, SVERRE. 2013. Neighborhood density in phonological alternations. *Berkeley Linguistics Society* 36, to appear.

STERIADE, DONCA. 1994. Positional neutralization. Los Angeles: University of California, Los Angeles, MS.

STERIADE, DONCA. 1999. Lexical conservatism in French adjectival liaison. *Formal perspectives in Romance linguistics*, ed. by Jean-Marc Authier, Barbara E. Bullock, and Lisa A. Reed, 243–70. Amsterdam: John Benjamins.

TESSIER, ANNE-MICHELLE. 2007. *Biases and stages in phonological acquisition*. Amherst: University of Massachusetts Amherst dissertation.

TOPORIŠIČ, JOŽE. 2000. *Slovenska slovnica*. Maribor: Obzorja.

TOPPING, DONALD. 1968. Chamorro vowel harmony. *Oceanic Linguistics* 7.67–79.

TRUBETZKOY, NIKOLAI. 1939. *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.

USSISHKIN, ADAM, and ANDREW WEDEL. 2009. Lexical access, effective contrast and patterns in the lexicon. *Perception in phonology*, ed. by Paul Boersma and Silke Hamann, 267–92. Berlin: Mouton de Gruyter.

VAN DE VIJVER, RUBEN, and DINAH BAER-HENNEY. 2011. Acquisition of voicing and vowel alternations in German. *Proceedings of the Boston University Conference on Language Development (BUCLD)* 35.603–15.

WALKER, RACHEL. 2012. *Vowel patterns in language*. Cambridge: Cambridge University Press.

WIESE, RICHARD. 1996. Phonological versus morphological rules: On German umlaut and ablaut. *Journal of Linguistics* 32.113–35.

WILSON, COLIN. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.945–82.

WOLF, MATTHEW. 2008. *Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model*. Amherst: University of Massachusetts Amherst dissertation.

YU, ALAN. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80.73–97.

ZHANG, JIE, and YUWEN LAI. 2008. Phonological knowledge beyond the lexicon in Taiwanese double reduplication. *Interfaces in Chinese phonology: Festschrift in honor of Matthew Y. Chen on his 70th birthday*, ed. by Yuchao E. Hsiao, Hui-chuan Hsu, Lian-hee Wee, and Dah-an Ho, 183–222. Taipei: Academia Sinica.

ZHANG, JIE; YUWEN LAI; and CRAIG SAILOR. 2011. Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua* 121.181–206.

ZIMMER, KARL. 1969. Psychological correlates of some Turkish morpheme structure conditions. *Language* 45.309–21.

ZURAW, KIE. 2000. *Patterned exceptions in phonology*. Los Angeles: University of California, Los Angeles dissertation.

Becker
Department of Linguistics
226 South College
University of Massachusetts
150 Hicks Way
Amherst, MA 01003
[michael@linguist.umass.edu]

Nevins
UCL Linguistics
Division of Psychology and Language Sciences
Chandler House
2 Wakefield Street
London WC1N 1PF, United Kingdom
[a.nevins@ucl.ac.uk]

Levine
Department of Linguistics
Harvard University
Boylston Hall 317
Cambridge, MA 02138
[jplevine@fas.harvard.edu]

[Received 17 April 2011;
accepted 21 September 2011]