Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000465

Contextual Variability and Exemplar Strength in Phonotactic Learning

Thomas Denby, Jeffrey Schecter, Sean Arn, Svetlin Dimov, and Matthew Goldrick
Northwestern University

Phonotactics—constraints on the position and combination of speech sounds within syllables—are subject to statistical differences that gradiently affect speaker and listener behavior (e.g., Vitevitch & Luce, 1999). What statistical properties drive the acquisition of such constraints? Because they are naturally highly correlated, previous work has been unable to dissociate the contribution of 2 properties: contextual variability (the number of unique phonological contexts in which a phonotactic pattern appears) and exemplar strength (the overall number of times the pattern appears). Using an artificial language learning paradigm, 3 experiments disentangled the effects of variability and strength, indexed by type and token frequency, respectively, on the learning of gradient phonotactics. When the 2 factors were decorrelated (Experiment 2), participants showed greater generalization of patterns advantaged for contextual variability, but not those advantaged for exemplar strength. When the 2 factors were anticorrelated (Experiment 3), participants preferred patterns advantaged in contextual variability, even though they were disadvantaged for exemplar strength. These results suggest that contextual variability is the key force driving phonotactic learning, as it allows learners to home in on the invariant features of the input.

Keywords: phonotactics, phonotactic learning, statistical learning, generalization, adult language learning

We effortlessly extend patterns over linguistic elements to novel structures. For example, after learning words like *king* and *him*, we can easily process nonsense words like *hing* but have difficulty hearing and producing nonsense syllables like *ngih*. The internalized representation of these sound structure patterns are referred to as *phonotactic constraints*. For example, /h/ is categorically constrained to the beginning of syllables in English, while /ŋ/ is categorically restricted to syllable-final position. Beyond categorical presence/absence, some phonological structures occur more often than others (e.g., syllable-final /s/ > syllable-final /z/ in English). As reviewed below, such gradient frequency differences result in gradient behavioral effects, with learning, perception, and production differing for high- versus low-frequency sound sequences.

What statistical properties drive the acquisition of such constraints, allowing us to generalize phonotactic patterns to novel forms? Associative models, such as the Neighborhood Activation Model (Vitevitch & Luce, 1999), connectionist models (e.g., Warker & Dell, 2006), and exemplar models (e.g., MINERVA 2; Goldinger, 1998), posit that the relevant factor is *exemplar strength*: the frequency with which a learner is exposed to a sound structure, as indexed by token frequency. In these associative models, an association (e.g., /s/ appears in syllable-final position) is strengthened with each exposure to items containing this structure. Under this account, the gradient preference for high token frequency /s/ versus low token frequency /z/ in syllable-final position reflects the differing strengths of those associations. The associative account makes a strong prediction that exemplar strength drives generalization: the degree to which learners generalize a sound structure to novel items should reflect the associative strength of that structure.

Other theories have assumed that learning is driven by contextual variability: the range of unique lexical, syllabic, or other phonological contexts in which a sound sequence appears (indexed by type frequency; e.g., Pierrehumbert, 2001; in morphology, see Bybee, 1988; Bybee, 1995; Bybee & Hopper, 2001). For example, syllable-final /s/ is preceded by a greater variety of vowels and vowel-onset pairs than syllable-final /z/. Such contextual variability may enhance phonotactic pattern learning by drawing the learner's attention to invariant features of the input—that is, the aspect of the input shared by a large number of distinct input syllables (e.g., syllable-final /s/).

Distinguishing these accounts has been difficult, as contextual variability and exemplar strength are highly correlated in natural language. All else being equal, a structure that occurs in many contexts will also occur over many token examples (see, e.g., Tamaoka & Makioka, 2004, for a quantitative analysis of such correlations in Japanese; but cf. Whalen, Giulivi, Nam, Levitt, Hallé, & Goldstein, 2012, for divergence in English spontaneous speech).

Despite the correlation of exemplar strength and contextual variability, the two factors may not equally contribute to learning. This study utilizes a series of artificial language experiments to decorrelate type and token frequency, allowing us to isolate the effects of contextual variability and exemplar strength on phonotactic pattern learning and examine their interaction. Distinguishing between the effects of these two factors should illuminate the mechanisms underlying phonotactic learning. We begin by review-

Thomas Denby, Jeffrey Schecter, Sean Arn, Svetlin Dimov, and Matthew Goldrick, Department of Linguistics, Northwestern University.

Correspondence concerning this article should be addressed to Thomas Denby, Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60657. E-mail: tdenby@u.northwestern.edu

ing previous research on gradient phonotactics, and then focus on the role of contextual variability and exemplar strength in phonotactic learning. We then present three experiments that decorrelate their effects on gradient phonotactic learning, and find that contextual variability is the most critical factor.

Gradient Phonotactics in Language Processing

In both development and adulthood, gradient phonotactic constraints influence how we process language. Infants listen longer to nonwords with high-frequency phonotactics over those with low-frequency phonotactics (Jusczyk & Luce, 1994); toddlers produce words with high-frequency phonotactics with greater accuracy than those with low-frequency phonotactics (Beckman & Edwards, 2000); adults judge nonwords with high-frequency phonotactics as more English-like than those with low-frequency phonotactics (Treiman, Kessler, Knewasser, Tincoff, & Bowman, 2000); nonwords with high-frequency phonotactics result in faster reaction times in discrimination and shadowing tasks for adults (Vitevitch & Luce, 1998, 1999), as well as picture naming (Vitevitch, Armbrüster, & Chu, 2004); adults are more likely to make speech errors on low-frequency structures, resulting in the production of high-frequency structures (Kupin, 1982; see Goldrick, 2011a, for a review); and experimentally learned frequency differences modulate the rates at which adult speakers' speech errors maintain syllable position (Goldrick & Larson, 2008).

Factors Facilitating the Acquisition of Phonotactics

What allows learners to acquire gradient phonotactics? A number of artificial language learning studies have suggested that acoustic variability can enhance learning. Seidl, Onishi, and Cristia (2014) found that infants acquired categorical phonotactic constraints when exposed to multiple talkers, but not single talkers. Richtsmeier, Gerken, and Ohala (2011) found similar results for gradient phonotactic constraints. Four-year-old children were trained to produce novel words in which the frequency of medial consonant sequences was varied. Production of the medial sequences improved when children heard the sequence in a variety of phonological contexts, spoken by multiple talkers. In contrast, no improvement was found when a single type was repeated by multiple talkers, or when multiple types were spoken by a single talker. This suggests that acquisition of gradient phonotactics is facilitated by acoustic variability when accompanied by contextual variability.

While acoustic variability appears to facilitate learning, the relative contributions of contextual variability and exemplar strength have been less clear. Some studies have focused on naturally occurring examples where these two factors are decorrelated (e.g., word-initial /ð/ in English appears in very few words, but has very high token frequency, as it appears in high token frequency function words like "the" and "that"). Results from perception (Pierrehumbert, 2001) and production (Buchwald, 2005) have been argued to show little sensitivity to exemplar strength compared with contextual variability. Similarly, computational models of phonotactic learning have shown better performance when using type, rather than token, statistics as the basis for training corpora (Hayes & Wilson, 2008). However, the correlation of token and type frequency naturally limits the power of such analyses.

An alternative approach, which we pursue here, is to use artificial language learning to create circumstances in which these two factors are decorrelated. This builds on the work of Richtsmeier (2011), who used word-likeness judgments to assess adults' learning of novel gradient phonotactic constraints. Richtsmeier (2011) exposed participants to advantaged and disadvantaged sound structures. This advantage was, depending on the condition, cued by high levels of talker variability, contextual variability, and/or exemplar strength. Because of the nature of the design, it is unclear which factor drives learning; in each condition, multiple factors were covaried. Critically, in the condition meant to isolate the influence of contextual variability, the advantaged structure not only appeared in more unique contexts, but also had higher exemplar strength (as each of these unique contexts was repeated the same number of times for each structure, resulting in a greater number of total instances of the pattern). Given the confounding of these factors, it is unclear how they affect learning of gradient phonotactics.

Overview of Experiments

In this study, we utilized an artificial language learning paradigm to investigate the acquisition of gradient phonotactic constraints. Participants were exposed to a series of syllables reflecting a gradient constraint (e.g., during familiarization /f/ is more frequent in coda than /s/). After participants were familiarized with syllables from these two phonotactic patterns, a few novel generalization syllables, each of which appeared once, were presented, intermixed with more repetitions of familiarization syllables. Generalization syllables equally reflected both phonotactic patterns that appeared in familiarization syllables (e.g., coda-/f/ and coda-/s/ were equally frequent in generalization syllables). Moreover, stimuli were counterbalanced such that phonological overlap between each generalization syllable and the entire set of familiarization syllables was controlled.

After each syllable was presented, participants indicated whether or not they had previously heard that syllable during the experiment. After seeing multiple repetitions of familiarization syllables, participants were very likely to respond "yes" on these syllables. The crucial measure was how often participants responded "yes" on generalization syllables, as these syllables had not yet been experienced by participants. If participants are internalizing a phonotactic pattern they are being exposed to, we expect them to incorrectly believe they had previously heard syllables reflecting this pattern. Similarly, if participants are not internalizing a pattern, we expect them to reject previously seen syllables reflecting this pattern. More important, false recognition rates for generalization syllables are always assessed relative to a within-participant baseline. For example, to isolate the effects of frequency, the false recognition rate for generalization syllables reflecting a high-frequency pattern (e.g., coda /f/) is compared with the false recognition rate of generalization syllables reflecting a low-frequency pattern (e.g., coda /s/). If frequency affects phonotactic learning, participants should be more likely to false alarm on generalization syllables reflecting the high-frequency pattern than those reflecting the low-frequency pattern.

Previous investigations using this paradigm have shown it to be reliable, with participants able to learn categorical phonotactic constraints restricting consonants to a particular syllable position

(Bernard, 2015). Steele, Denby, Chan, and Goldrick (2015) showed that this paradigm can be successfully deployed online. Participants exposed to constraints over the web can successfully acquire categorical constraints on coda consonants conditioned on the preceding vowel. These previous results suggest a number of advantages for this approach over other paradigms frequently used to measure participants' phonotactic knowledge. It is a relatively implicit task—participants are not asked to make explicit metalinguistic judgments, unlike in word-likeness or acceptability tasks (e.g., Richtsmeier, 2011; see Goldrick, 2011b, for discussion of potential issues with metalinguistic tasks). Additionally, perception paradigms that can be utilized over the web can be used to gather large amounts of data quickly, allowing experimenters to contrast a number of different exposure conditions. In contrast, production paradigms examining the distribution of speech errors (e.g., Warker & Dell, 2006), require very labor intensive data analysis, placing practical barriers to collecting large amounts of data.

In this study, we use this implicit paradigm to disentangle the effects of contextual variability and exemplar strength on learning by varying the type and token statistics of familiarization syllables across experiments. As summarized in Table 1, this allowed us to explore how each factor individually contributes to learning (Experiment 2), as well as how they interact with one another (Experiment 3).

In Experiment 1A, we verify that the artificial language learning paradigm that has previously been used to examine the learning of categorical phonotactic constraints (Bernard, 2015; Steele et al., 2015) can be extended to gradient constraints. Then, in Experiment 1B, we examine if such constraints can be acquired in an online version of the paradigm. This facilitates recruitment of participants for Experiments 2A–3C, in which we decorrelate contextual variability and exemplar strength.

Overview of Experiment Design

All stimuli were consonant-vowel-consonant (CVC) monosyllables, made up of four consonants (/b, s, f, and n/) used as onsets and codas, and four vowels (/i, u, æ, and ɛ/) used as syllable nuclei. The result was a total of 64 possible syllables (4 consonants * 4 vowels * 4 consonants = 64).

Participants were exposed to two different syllable coda patterns; one of these coda patterns was advantaged for contextual variability, exemplar strength, or both (see Table 1), while the other was disadvantaged. The 64 syllables were equally divided into two patterns, each consisting of 32 syllables, based on their coda consonants: one phonotactic pattern consisted of syllables ending in /n/ or /f/, while the second phonotactic pattern consisted of syllables ending in /b/ or /s/ (see Figure 1 for diagram of counterbalancing). Arbitrary sets of coda consonants were used to avoid natural class biases (see Chambers, Onishi, & Fisher, 2010). These syllables were then equally divided again into two sets, each consisting of 16 syllables: one set served as familiarization syllables, while the other set served as generalization stimuli. Set A consisted of onset-vowel pairs /bɛ, bi, fæ, fi, nɛ, nu, sæ, su/ while Set B consisted of /bæ, bu, fɛ, fu, næ, ni, sɛ, si/. Both coda pattern and onset-vowel pairing was counterbalanced across participants (see Figure 2, e.g., stimulus sets).

The experiment was divided into six blocks (see Figure 3 for the design of Experiment 1) that were presented continuously. During the first two blocks—the familiarization phase—participants heard only the familiarization syllables, randomized within each block. The number of familiarization syllable types and tokens differed by experiment and pattern (i.e., advantaged vs. disadvantaged). In each of Blocks 3 through 6—the generalization phase—participants continued to hear the same set of randomized familiarization syllables. In addition, 32 total generalization syllables—eight per block, half reflecting the advantaged coda pattern and half reflecting the disadvantaged pattern—were randomly intermixed into generalization blocks. Each generalization syllable was only presented once.

Experiment 1

In Experiment 1, one phonotactic pattern was advantaged over the other pattern for both contextual variability and exemplar strength: the advantaged pattern contained four times as many unique syllables and four times as many total repetitions reflecting

Table 1
Contextual Variability and Exemplar Strength per Block (as Measured by Type and Token Frequencies, Respectively), and the Ratios between Advantaged and Disadvantaged Patterns, for All Experiments

	Block input statistics					
	Contextual variability (type frequency)			Exemplar strength (token frequency)		
Experiment	Pat A	Pat B	Ratio	Pat A	Pat B	Ratio
Correlated						
1A and 1B–Correlated	16	4	4:1	64	16	4:1
Isolated						
2A–Contextual variability	16	4	4:1	40	40	1:1
2B–Exemplar strength	16	16	1:1	64	16	4:1
Anticorrelated						
3A and 3B–Anticorrelated	16	4	4:1	16	64	4:1
3C–Short anticorrelated	8	4	2:1	16	32	2:1

Note. Pat = pattern.

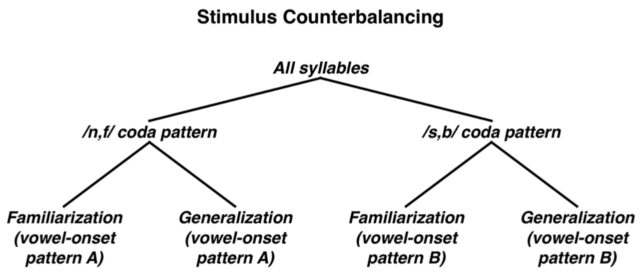

Figure 1. Diagram showing stimulus counterbalancing. There were 64 total syllables that are split into two equal groups of 32 based on coda pattern (/s,b/ vs. /n,f/). These groups are further split into groups of 16 syllables based on vowel-onset pattern: pattern A (onset vowel pairs *be, bi, fa, fi, ne, nu, sa, su*) versus pattern B (*ba, bu, fe, fu, na, ni, se, si*).

the phonotactic pattern as its disadvantaged counterpart. This simulates the statistics of gradient phonotactic constraints in natural language, as contextual variability and exemplar strength are highly correlated. Experiments 1A and 1B were identical in design, but the former was conducted in our lab, while the latter was conducted online.

The small set of familiarization syllables reflecting the disadvantaged pattern were counterbalanced such that they contained each onset and nucleus. As such, despite having fewer unique syllables, the degree of phonological overlap between the disadvantaged familiarization syllables and the generalization syllables was identical to that between the advantaged familiarization syllables and the generalization syllables with one exception: generalization syllables reflecting the disadvantaged pattern shared a coda with fewer familiarization syllables. For example, imagine the advantaged coda pattern was /s,b/ and the disadvantaged coda pattern was /n,f /, and participants are exposed to the generalization syllables *bes* and *suf*. Because *bes* follows the advantaged pattern, it shares a coda consonant with four times as many familiarization syllable types and tokens as *suf*. However, *bes* and *suf* share an equal number of onsets and nuclei with the set of familiarization syllables. In other words, outside of differences in coda consonant overlap because of advantaged versus disadvantaged patterns, no generalization syllables were closer or further to the familiarization set in phonological distance. Counterbalancing was done in this way for each pattern for all following experiments.

Experiment 1A: Variability/Strength Correlated, In-Lab

Method

Participants. Thirty-seven adult native English speakers with no history of speech or hearing impairments participated in the experiment for $7 compensation or course credit. The experiment took about 25 min on average.

Materials. Stimuli were recorded in a sound-attenuated booth by a male, native speaker of English and normalized to 60 dB SPL. The familiarization syllables consisted of 16 advantaged syllables types and 4 disadvantaged syllable types; all familiarization syllables were repeated four times per block. The result was 80 familiarization trials per block: 64 advantaged tokens per block (16 types * 4 repetitions) and 16 disadvantaged tokens per block (4 types * 4 repetitions = 16 tokens). In addition, 32 generalization syllables were included, equally spread across Blocks 3–6. The experiment had 512 total trials (80 familiarization tokens * 6 blocks + 32 generalization syllables). Crucially, the disadvantaged subset—that only included four syllable types—was counterbalanced such that it contained all four vowels and onsets, yielding 16 total lists (see Figure 2). List order was randomized for each participant within each block.

Procedure. Participants were seated in a sound-treated booth facing a computer screen and button box. Stimuli were presented over headphones. Instructions presented before the familiarization phase asked participants to push a button indicating whether or not that syllable had already been presented during the experiment. Stimuli were then played one at a time, with the next stimulus occurring following a participant response. Stimuli were played continuously, with the exception of a break at the experiment midpoint (between Blocks 3 and 4).

	EXPERIMENT 1A & 1B CORRELATED	EXPERIMENT 2A CV ISOLATED	EXPERIMENT 2B ES ISOLATED	EXPERIMENT 3A & 3B ANTI-CORRELATED	EXPERIMENT 3C SHORT ANTI-CORRELATED
Example Advantaged set	{bas, seb, nis, fub, bub, sis, nab, fes, bab, ses, nib, fus, bus, sib, nas, feb} *x 4 repetitions/block*	{bas, seb, nis, fub, bub, sis, nab, fes, bab, ses, nib, fus, bus, sib, nas, feb} *x 2 or 3 repetition/block*	{bas, seb, nis, fub, bub, sis, nab, fes, bab, ses, nib, fus, bus, sib, nas, feb} *x 4 repetitions/block*	{bas, seb, nis, fub, bub, sis, nab, fes, bab, ses, nib, fus, bus, sib, nas, feb} *x 1 repetitions/block*	{bis, fus, nub, sib, bib, fub, nus, sis} *x 2 repetitions/block*
Example Disadvantaged Set	{ban, sef, nin, fuf} *x 4 repetitions/block*	{ban, sef, nin, fuf} *x 10 repetitions/block*	{ban, sef, nin, fuf, buf, sin, naf, fen, baf, sen, nif, fun, bun, sif, nan, fef} *x 1 repetition/block*	{ban, sef, nin, fuf} *x 16 repetitions/block*	{bif, sin, nun, fuf} *x 8 repetitions/block*
Example Generalization set	bes, sab, nus, fib, bib, sus, neb, fas, beb, sas, nub, fis, bis, sub, nes, fab, ben, saf, nun, fif, bif, sun, nef, fan, bef, san, nuf, fin, bin, suf, nen, faf				bun, fin, nif, suf, buf, fif, nin, sun, bus, fis, nib, sub, bub, fib, nis, sus

Figure 2. Example stimulus sets for each experiment. Coda consonants conforming to the advantaged pattern are indicated with red; coda consonants conforming to the disadvantaged pattern are indicated with blue. Note that this figure represents a set of stimuli that one participant could be exposed to; it does not account for counterbalancing codas or onset/vowel pairs across participants. See the online article for the color version of this figure.

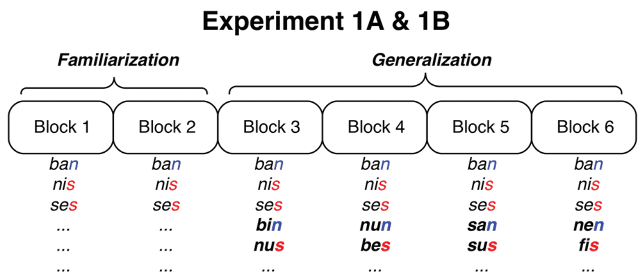

Figure 3. Design of Experiments 1A and 1B. The advantaged pattern (red) is repeated in a greater number of instances and appears in a greater variety of syllabic contexts than the disadvantaged pattern (blue). The same set of familiarization syllables is repeated in each of six blocks; in the final four blocks, participants are also exposed to single presentations of novel generalization syllables (bold), half reflecting each pattern. Note that only a fraction of all stimuli are shown here; stimuli not shown are denoted by ellipses. See the online article for the color version of this figure.

Results

To ensure participants were attending to the task, in generalization blocks (Blocks 3–6), participants had to respond "yes" on at least 90% of familiarization trials reflecting each pattern (i.e., stimuli they had heard before), and respond "no" on at least 10% of generalization trials (i.e., novel stimuli). If participants did not pass these criteria their data were excluded (although they were still compensated or given credit). Data from the remaining 32 participants (86.4%) was analyzed.

Analyses below are based on data from generalization blocks (i.e., Blocks 3–6); 95% confidence intervals (CI) for all analyses were estimated using a bootstrap method. This estimates the distribution of a statistic by repeatedly resampling from the observations with replacement. Here, 1,000 replicates were used to estimate the distribution of the mean across participants.

The mean "yes" response rate on familiarization syllables was extremely high: 95.1% (CI [94.6%, 95.6%]). The crucial measure that reflected learning of the pattern was generalization to novel syllables (i.e., participants mistakenly believing they had heard the novel syllable earlier in the experiment and responding with a false positive). As can be seen in Figure 4, participants false alarmed at a higher rate on novel syllables reflecting the variability/strength-advantaged pattern (mean: 58.5%) than for those reflecting the disadvantaged pattern (mean: 41.0%). The mean advantage for the advantaged pattern was 17.6% (CI [10.9%, 24.0%]).

These differences were statistically assessed via a logistic mixed-effects regression on the proportion of trials on which participants responded "yes" to generalization stimuli. There was one fixed contrast-coded effect: experimental condition (i.e., generalization stimuli reflecting the favored vs. disfavored pattern). Random slopes and intercepts for participants and syllable were included, along with a random slope for experimental condition by participant and syllable. Significance was assessed using a χ^2 model comparison (Barr, Levy, Scheepers, & Tily, 2013). This analysis confirmed that participants were significantly more likely to respond "yes" to generalization syllables for the advantaged pattern than the disadvantaged pattern ($\beta = .89$, $SE\ \beta = 0.17$, $\chi^2(1) = 20.02$, $p < .0001$).

We then considered the possibility that these effects may vary over the course of the experiment. Intuitively, given that "yes" is the correct response for the vast majority of stimuli, as more and more stimuli are encountered participants should increase their bias toward responding "yes." Therefore, we extended the regression model above to include block (centered) as a main effect, allowing it to interact with pattern type (including random intercepts for each effect; all models including block used this effects structure). There was a marginal main effect for block ($\beta = 0.16$, $SE\ \beta = 0.09$, $\chi^2(1) = 3.11$, $p = .08$); the interaction between block and condition was not significant ($\beta = 0.00$, $SE\ \beta = 0.14$, $\chi^2(1) = 0.00$, $p = .99$). In other words, participants may be more likely to respond "yes" as the experiment continues, but this effect is not significantly different for advantaged or disadvantaged patterns.

Experiment 1B: Variability/Strength Correlated, Online

Experiment 1A extended the results of the artificial language learning paradigm to show that participants can acquire gradient phonotactic constraints. Steele et al. (2015) showed that this paradigm could be used online to study the acquisition of categorical constraints. In Experiment 1B, we examine whether gradient constraints can also be acquired outside of the laboratory.

Method

Participants. Fifty-four participants were recruited using Amazon Mechanical Turk (AMT; Buhrmester, Kwang, & Gosling, 2011). Participants were paid $3 each, which, given the short time to complete the experiment (~15 min), is roughly equivalent to hourly rate used in our lab ($10/hr). The experiment was restricted to workers with IP addresses in the United States. Of the 54 participants, 10 were excluded because of server error, in which one or more trial was not recorded. Out of the remaining 44 participants, analysis focused on the 32 participants (72.7%) that passed the exclusion criteria outlined above.

All participants who passed the criteria self-identified as having no language impairments; one participant self-identified as a non-native speaker of English. Results were not qualitatively affected by the exclusion of that participant's data, and as such that par-

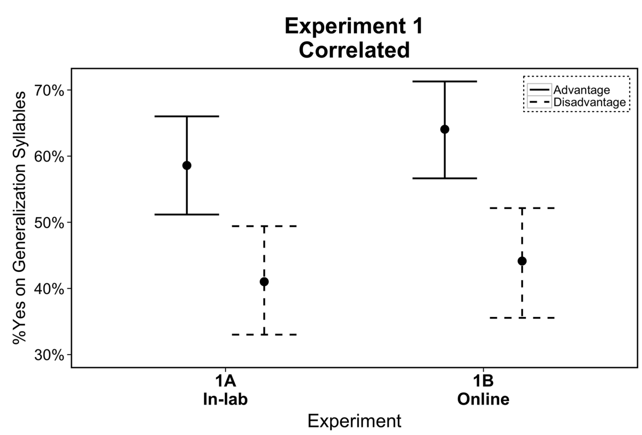

Figure 4. Proportion "yes" responses to generalization syllables for variability/strength-advantaged and disadvantaged patterns for Experiments 1A and 1B. Error bars reflect bootstrapped 95% confidence interval.

ticipant's data was not excluded. (The same procedure was utilized for participants in all experiments who reported being a nonnative speaker or having a language impairment).

Materials. Stimuli were identical to Experiment 1A.

Procedure. Participants were given a pre-experiment audio test to ensure their audio was working, and their computer volume was at a comfortable level. The audio test was a two-stage process. In both stages, participants had to listen to a word of English and correctly type it. The recordings were made by a different speaker than the one who recorded the experimental stimuli. Participants were allowed to play the recordings as many times as necessary.

As in Experiment 1A, participants performed a recognition memory task. The question "Have you heard this syllable before?" was on the screen for the entire experiment. On each trial, an auditory stimulus was presented. Participants answered the question by clicking a "Yes" or "No" button on the screen. There was a 500 ms interstimulus interval, during which the "Yes" and "No" buttons disappeared. Participants had unlimited time to answer the question, and no feedback was provided. Unlike Experiment 1A, there were no breaks in between experimental blocks. This procedure was utilized in all subsequent experiments reported below.

Results

Analysis methods followed Experiment 1A. The mean "yes" response rate on familiarization syllables was 94.1% (CI [93.1%, 94.9%]). As can be seen in Figure 4, participants responded "yes" at a higher rate on novel syllables reflecting variability/strength-advantaged pattern (mean: 64.1%) than for those reflecting the disadvantaged pattern (mean: 44.1%). The mean difference in "yes" response rate for the advantaged pattern was 19.9% (CI [14.1%, 26.7%]). A logistic mixed-effects regression was run, identical to that in Experiment 1A, with subjects and items treated as random effects. The results revealed that the difference between conditions was significant ($\beta = 1.07$, $SE\ \beta = 0.19$, $\chi^2(1) = 23.75$, $p < .0001$). The subsequent block analysis, however, showed no significant main effect for block ($\beta = 0.09$, $SE\ \beta = 0.07$, $\chi^2(1) = 1.75$, $p = .19$) and no significant interaction between block and condition ($\beta = 0.13$, $SE\ \beta = 0.13$, $\chi^2(1) = 0.88$, $p = .34$).

To investigate whether response rates differed based on whether participants took the experiment in the lab or online (i.e., Experiment 1A or 1B), the logistic mixed effects regression models used above were extended to include a contrast-coded factor for experiment (along with a random slope for experiment by participant and syllable). This interacted with pattern advantage (i.e., advantaged vs. disadvantaged patterns), including a random slope for the interaction by syllable. Results revealed that there was no significant interaction between experiment and pattern advantage ($\beta = 0.14$, $SE\ \beta = 0.24$, $\chi^2(1) = 0.33$, $p = .56$), suggesting that the setting of the experiment did not influence the degree to which participants generalized the novel phonotactic constraint. In addition, there was no main effect for experiment ($\chi^2(1) = 0.69$, $p = .40$).

Discussion

Both in the lab and online, participants generalize a gradient phonotactic pattern that is advantaged in both contextual variability and exemplar strength to novel syllables. In subsequent experiments, we examine the relative contribution of these two factors to the learning of constraints.

Experiment 2: Variability, Strength Isolated

In Experiment 2, two individual factors associated with gradient phonotactic constraints were isolated to investigate their contributions to generalization: contextual variability (Experiment 2A) and exemplar strength (Experiment 2B). As noted in the introduction, these factors are highly correlated in natural language, making it difficult to distinguish between their effects with more naturalistic methods. Artificial language experiments are particularly useful for this type of investigation, as it allows us to decorrelate and manipulate different sources of information that are usually copresent in the input.

Experiment 2A: Contextual Variability Isolated

Experiment 2A isolated contextual variability: codas in the advantaged pattern appeared in a greater number of unique syllables (i.e., onset/vowel contexts) than those in the disadvantaged pattern, while exemplar strength was held constant. This contextual variability may, in turn, highlight the invariant features of the input (in this case, the recurring coda consonants). If such a mechanism underlies generalization, we would predict the variability-advantaged pattern to be generalized to novel syllables at a greater rate than its disadvantaged counterpart.

Method

Participants. There were 57 participants that were recruited using AMT. Data from 10 participants was excluded because of data loss from server error; of the remaining 47, 33 passed the experimental criteria (70.2%). Because of a software error, one stimulus list had two participants who both passed the criteria; the second, "duplicate" participant that completed the task was excluded. (The same procedure was used for all subsequent experiments.) Data from the remaining 32 participants (all self-identified native English speakers with no language impairments) is analyzed below.

Materials. Familiarization and generalization syllables were defined as in Experiment 1. Within the familiarization set, the variability-advantaged pattern consisted of 16 unique syllables, half of which were repeated twice per block, and half three times per block, for a total of 40 tokens per block (16 syllables * 2 or 3 repetitions; see Figure 2). The disadvantaged pattern consisted of subset of 4 syllables from the set of 16 possible syllables (counterbalanced as in Experiment 1), each of which was repeated 10 times per block. Thus, both patterns had an identical number of tokens, but the variability-advantaged pattern had four times the number of types. There were 80 tokens in each of the six blocks (40 from each pattern), along with 32 generalization syllables, for a total of 512 total trials.

Results

The mean "yes" response rate on familiarization syllables was 94.6% (CI [94.0%, 95.3%]). For generalization syllables, as can be seen in Figure 5, participants responded "yes" at a higher rate on novel syllables reflecting the variability-advantaged pattern (mean:

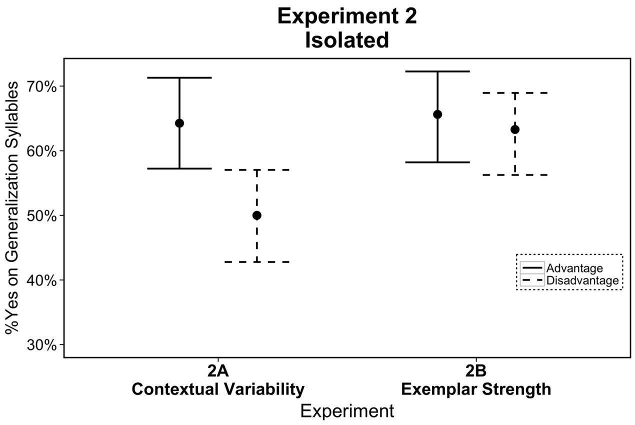

Figure 5. Proportion "yes" responses to generalization syllables for Experiments 2A (variability advantaged) and 2B (strength advantaged). Error bars reflect bootstrapped 95% confidence interval.

64.3%) than for those reflecting the disadvantaged pattern (mean: 50.0%). The mean difference between the variability-advantaged and disadvantaged patterns was 14.3% (CI [9.4%, 19.5%]). A logistic mixed-effects regression, identical to those in previous experiments, revealed that the difference between conditions was significant ($\beta = 0.75$, *SE* $\beta = 0.15$, $\chi^2(1) = 21.92$, $p < .0001$). A subsequent block analysis also revealed a significant main effect for block ($\beta = 0.23$, *SE* $\beta = 0.07$, $\chi^2(1) = 12.85$, $p < .01$), but no significant interaction between block and condition ($\beta = -0.03$, *SE* $\beta = 0.13$, $\chi^2(1) = 0.06$, $p = .80$).

Experiment 2B: Exemplar Strength Isolated

Experiment 2B isolated exemplar strength and held contextual variability constant: all coda consonants appeared in the same number of unique syllables, but in the advantaged pattern, those syllables were repeated four times as frequently. If the strength of the individual items making up a given phonotactic pattern positively contributes to that pattern's overall strength or generalizability, then we predict that high strength should result in increased generalization.

Method

Participants. There were 83 participants who were recruited using AMT. Data from 14 participants was excluded due data loss from server error; of the remaining 69, 33 passed the criteria (47.8%). The data from one duplicate participant was excluded. All participants who passed the criteria self-identified as native English speakers with no language impairments.

Materials. Familiarization and generalization syllables were defined as in Experiment 1. Within the familiarization set, both the strength-advantaged and disadvantaged pattern consisted of 16 unique syllables (see Figure 2). Each syllable in the strength-advantaged pattern was repeated four times per block versus once per block for each syllable in the disadvantaged pattern. Thus, both patterns had an identical number of types, but the strength-advantaged pattern had four times the number of tokens. There were 80 familiarization items in each of the six blocks (64 strength-advantaged + 16 disadvantaged), along with 32 generalization syllables, for a total of 512 total trials.

Results

The mean "yes" response rate on familiarization syllables was 92.6% (CI [91.9%, 93.2%]). As can be seen in Figure 5, participants responded "yes" at a similar rate on novel syllables reflecting the strength-advantaged pattern (mean: 65.6%) and those reflecting the disadvantaged pattern (mean: 63.28%). The mean advantage for the strength-advantaged pattern was 2.3% (CI [−3.3%, 8.2%]). A logistic mixed-effects regression and χ^2 model comparison revealed that the difference between conditions was not significant ($\beta = 0.09$, *SE* $\beta = 0.17$, $\chi^2(1) = 0.3$, $p = .58$). The subsequent block analysis showed there was, however, a significant main effect for block ($\beta = 0.38$, *SE* $\beta = 0.07$, $\chi^2(1) = 30.32$, $p < .0001$), suggesting that, similarly to Experiment 2A, participants were more likely to respond "yes" in later blocks. Moreover, there was a significant interaction between block and condition ($\beta = .29$, *SE* $\beta = 0.14$, $\chi^2(1) = 4.2$, $p < .05$), as participants' preference for the advantaged pattern increased slightly over the course of the experiment.

Discussion

Experiment 2 isolated contextual variability and exemplar strength and measured their effect on generalization. Our results suggest that contextual variability has a stronger effect on generalization than exemplar strength. To confirm this finding, we ran a logistic mixed-effects regression, comparing results from Experiments 2A and 2B. This paralleled the structure of the model used to compare Experiments 1A and 1B, except that random slopes and intercepts were decorrelated for subjects because of convergence issues. A significant interaction between pattern advantage and experiment was found ($\beta = 0.58$, *SE* $\beta = 0.20$, $\chi^2(1) = 8.18$, $p < .01$), suggesting that participants generalize more because of high contextual variability than exemplar strength. (There was no main effect of experiment; $\chi^2(1) = 2.39$, $p = .12$).

Experiment 3: Variability, Strength Anticorrelated; Acoustically Variable Stimuli

Experiment 2 suggests that exemplar strength does not, in isolation, have a significant influence on the learning of gradient phonotactics; that said, it may still modulate generalization through interaction with other factors. In Experiment 3A, we examine the interaction between variability and strength by anticorrelating the two factors: participants were exposed to one pattern with high contextual variability, and another with high exemplar strength within the same experiment.

One possible confound of Experiment 3A (as well as 2B) is that the effect of increased exemplar strength is eliminated by using acoustically identical repetitions. To address this confound, acoustic variability, in the form of duration variation, was added to both patterns in Experiment 3B.

Finally, Experiment 3C examines whether the preceding results are driven by the specific 4:1 advantage in either contextual variability or exemplar strength utilized in these experiments. This experiment anticorrelates strength and contextual variability, but cuts the relative frequency advantages in half (while including acoustic variability). This design also allows us to examine whether the preceding results are driven by the degree of overlap between familiarization and generalization syllables (see below for further discussion).

Experiment 3A: Variability, Strength Anticorrelated

In Experiment 3A, contextual variability and exemplar strength were anticorrelated: coda consonants in the variability-advantaged pattern appeared in many syllables, but with few overall repetitions. In the strength-advantaged pattern, on the other hand, coda consonants appeared in only a few unique syllables, but with many repetitions.

Participants

There were 85 participants who were recruited using AMT. Data from 13 participants was excluded because of data loss from server error; of the remaining 72, 34 passed the criteria (47.2%). The data from two duplicate participants was excluded. All participants who passed the criteria self-identified as having no language impairments; two participants self-identified as nonnative speakers.

Method

Materials. Division of syllables into familiarization and generalization followed the experiments above. The contextual variability-advantaged pattern consisted of 16 unique syllables, each repeated once per block. The strength-advantaged pattern consisted of just four unique syllables, each repeated 16 times per block (see Figure 2). This resulted in the strength-advantaged pattern having a 4:1 token frequency ratio, and, similarly, the contextual variability-favored pattern having a 4:1 type frequency ratio. As in previous experiments, 32 novel generalization syllables were included in Blocks 3–6, resulting in 512 total trials.

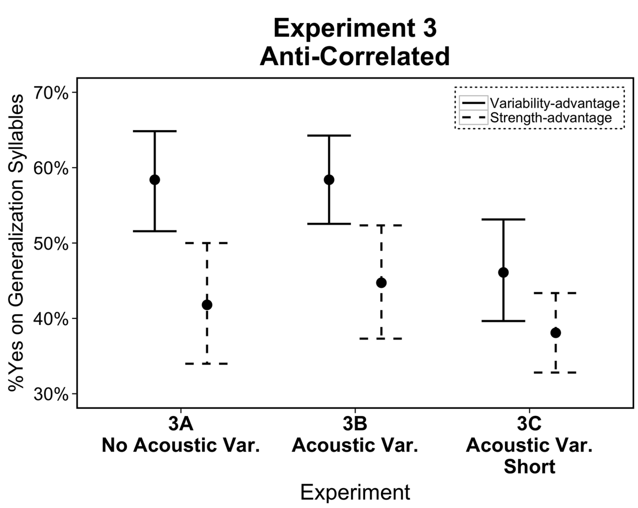

Figure 6. Proportion "yes" responses to generalization syllables for contextual variability-advantaged and strength-advantaged patterns for Experiments 3A (anticorrelated), 3B (anticorrelated with acoustic variability), and 3C (anticorrelated with duration variability and reduced pattern ratio). Error bars reflect bootstrapped 95% confidence interval.

Results

The mean "yes" response rate on familiarization syllables was 95.2% (CI [94.8%, 95.6%]). As can be seen in Figure 6, participants responded "yes" at a higher rate on novel generalization syllables reflecting the contextual variability-advantaged pattern (mean: 58.4%) than for those reflecting the strength-advantaged pattern (mean: 41.8%). The mean advantage for the contextual variability-advantaged pattern was 16.6% (CI [10.4%, 22.7%]). A logistic mixed-effects regression, identical to those in previous experiments revealed that the difference between conditions was significant ($\beta = 0.82$, $SE\ \beta = 0.15$, $\chi^2(1) = 22.2$, $p < .0001$). Consistent with Experiment 2, this suggests that contextual variability is the key factor in learning gradient phonotactics.

The subsequent block analysis revealed a significant main effect for block ($\beta = 0.29$, $SE\ \beta = 0.06$, $\chi^2(1) = 20.0$, $p < .0001$), as well a marginal interaction between block and condition ($\beta = 0.23$, $SE\ \beta = 0.13$, $\chi^2(1) = 3.12$, $p = .08$).

Experiment 3B: Variability, Strength Anticorrelated; Acoustically Variable Stimuli

One possible confound of Experiment 3A is that the contextual variability-advantaged pattern contains a broader variety of unique syllables and, therefore, by definition, exhibits more acoustic variability than a smaller number of unique syllables repeated more frequently (the strength-favored pattern). In addition, it is possible that there is an interaction between exemplar strength and acoustic variability: perhaps the two factors can only modulate generalization in combination (see Richtsmeier et al., 2011). Including acoustic variability also provides a more naturalistic test of learning, as acoustic variability and exemplar strength are necessarily correlated in natural speech (i.e., no two utterances, even of segmentally identical material, are acoustically identical). To address this confound, acoustic variability was added to both patterns in Experiment 3B.

Duration variability was chosen as an appropriate form of acoustic variability, as duration is a crucial phonetic cue (e.g., voice onset time, vowel length distinctions, etc.). In addition, there is evidence that listeners track duration variability during learning: Sommers and Barcroft (2007), for example, found that duration variability aids second language vocabulary learning, suggesting that participants interpret duration variability as linguistically meaningful and are able to attend to it in an experimental setting.

Participants

There were 82 participants who were recruited using AMT. Data from 1 participant was excluded because of data loss from server error; of the remaining 81 participants, 33 passed the criteria (40.7%). All participants who passed the criteria self-identified as having no language impairments; one participant self-identified as a nonnative speaker.

Method

Materials. To instantiate acoustic variation, duration was manipulated across the entire sound file using Praat (Boersma & Weenink, 2012), ranging from 70–130% duration of the baseline stimulus (Sommers & Barcroft, 2007, manipulated duration to a similar degree, suggesting such differences are salient to listeners). From the baseline, duration was increased in eight steps and decreased in eight steps, resulting in 17 versions of each stimulus differing in duration (eight longer, eight shorter, and the baseline stimulus). For both the shorter and longer stimuli, six steps were in increments of 4% and two steps were in increments of 3%. Because of experimenter error, there was no increase in length for Steps 5 and 6 of the stimuli manipulated to be longer, meaning Steps 4, 5, and 6 of the longer stimuli were all equal in length.

The distribution of syllables across conditions followed Experiment 3A. All generalization stimuli were presented at the baseline duration. In each block, each of the 16 stimuli in the contextual variability-advantaged pattern was played at one of each of the 16 manipulated durations. The stimulus-duration combination was switched each block, such that each stimulus was played at a different duration in each block. For the strength-favored pattern, each of the 16 repetitions of each syllable per block was played at a different manipulated duration.

Results

The mean "yes" response rate on familiarization syllables was 95.1% (CI [94.8%, 95.4%]). As can be seen in Figure 6, participants responded "yes" at a higher rate on novel generalization syllables reflecting the contextual variability-advantaged pattern (mean: 58.4%) than for those reflecting the strength-advantaged pattern (mean: 44.7%). The mean advantage for the contextual variability-favored pattern was 13.7% (CI [6.4%, 20.3%]). A logistic mixed-effects regression revealed that the difference between conditions was significant ($\beta = 0.68$, $SE\ \beta = 0.17$, $\chi^2(1) = 12.81$, $p < .0001$). This suggests that the results of Experiment 3A are not simply a result of using acoustically identical stimuli. In addition, the subsequent block analysis revealed a marginally significant main effect for block ($\beta = 0.12$, $SE\ \beta = 0.06$, $\chi^2(1) = 3.67$, $p = .06$), but no interaction between block and condition ($\beta = -0.03$, $SE\ \beta = 0.13$, $\chi^2(1) = 0.04$, $p = .84$).

Experiment 3C: Anticorrelated, Acoustically Variable Stimuli; Reduced Pattern Ratio

Previous experiments in this study examined a very narrow space of the possible input statistics: patterns with a 4:1 advantage in either exemplar strength or contextual variability. In an effort to explore more of the space of possible input statistics, Experiment 3C anticorrelates strength and contextual variability, but cuts the relative frequency advantages in half. That is, patterns have a 2:1 advantage in either type or token frequency (as such, the experiment has many fewer total trials). In addition to widening the space of input statistics, this is also a more stringent test of the increased influence of contextual variability. Is a 2:1 ratio large enough for the variability-advantaged pattern to show a higher rate of generalization? Note that Experiment 3C, like Experiment 3B, was manipulated for increased acoustic variability across both patterns.

This design also allows us to control the overlap between familiarization and generalization syllables. In Experiments 3A and 3B, there are only 4 distinct syllables in the familiarization set for the strength-advantaged pattern. Because there are eight possible combinations of vowels and coda consonants (rimes) for each pattern, this design requires participants to generalize to novel rimes to successfully acquire the strength-advantaged pattern. Rimes are well-known to be highly salient for English speakers (Kessler & Treiman, 1997); the lack of overlap could inhibit learning. In Experiment 3C, there are only four possible rimes (two vowels and two codas). Participants are exposed to all of these during familiarization, removing this potential barrier to learning the strength-advantaged pattern.

Participants

As there were half as many generalization items per participant relative to preceding experiments, the number of participants was increased in an effort to maintain statistical power. There were 104 participants who were recruited using AMT. Data from 3 participants was excluded because of data loss from server error; of the remaining 101 participants, 65 passed the criteria (64.4%). The data from one duplicate participant was excluded. Three participants who passed the criteria self-identified as nonnative speakers of English; all participants self-identified as having no language impairments.

Materials

Stimuli consisted of one half of the 64 syllables used in previous experiments—the same four consonants (/b, s, f, and n/) were used as onsets and codas, but were combined with only two vowels (/i, u/), as opposed to the four used in earlier experiments (see Figure 2). Counterbalancing was similar to Experiments 3A and 3B. Syllables were split into groups by coda pattern and onset/vowel combination, resulting in only eight syllables for familiarization versus generalization sets. These groups were then split in half for the strength-advantaged pattern, with each of these subgroups of four syllables balanced for onset, vowel, and coda.

Within each familiarization block, the eight contextual variability-advantaged syllables were each presented twice, while each of the four strength-advantaged syllables was presented eight times. This resulted in the strength-advantaged pattern having a 2:1 token frequency ratio, and the contextual variability-advantaged pattern having a 2:1 type frequency ratio.

There were 16, as opposed to 32, generalization syllables: 8 each from both the contextual variability- and strength-advantaged patterns. As in previous experiments, generalization syllables were equally distributed across the four latter blocks. Thus, there were four generalization syllables—two from each coda pattern—presented in each of the final four blocks. The experiment had 304 total trials.

A subset of the duration-manipulated stimuli from Experiment 3B were used: from the baseline, duration was increased and decreased in four, rather than eight, steps. Thus, there were nine versions of each stimulus differing in duration (four longer, four shorter, and the baseline stimulus). The range of duration variability was identical to that in Experiment 3B, but with fewer unique steps utilized along the continuum.

Results

The mean "yes" response rate on familiarization syllables was 94.3% (CI [93.8, 94.8%]). As can be seen in Figure 6, participants responded "yes" at a higher rate on novel syllables reflecting the contextual variability-advantaged pattern (mean: 46.1%) than for those reflecting the strength-advantaged pattern (mean: 38.1%). The mean advantage for the contextual variability-favored pattern was 8.0% (CI [2.3%, 13.3%]). A logistic mixed-effects regression revealed that the difference between conditions was significant ($\beta = 0.37$, $SE\ \beta = 0.17$, $\chi^2(1) = 4.13$, $p < .05$). In the subsequent block analysis, no main effect was found for block ($\beta = 0.09$, $SE\ \beta = 0.06$, $\chi^2(1) = 2.04$, $p = .15$) and no interaction between block and condition was found ($\beta = -0.05$, $SE\ \beta = 0.13$, $\chi^2(1) = 0.14$, $p = .71$).

Discussion

Results from Experiment 3 consistently show that patterns with high contextual variability are generalized more than those with high exemplar strength. The addition of acoustic variability in Experiment 3B does not change this outcome. Furthermore, in Experiment 3C, the reduction of contextual variability and strength advantages, along with an increase in overlap between familiarization and generalization syllables for the strength-advantaged pattern, did not alter the advantage for contextual variability.

To confirm these findings, two mixed-effects logistic regressions were run, comparing "yes" response rates across pairs of experiments. The interaction between experiment and pattern advantage for Experiments 3A and 3B was not significant ($\beta = 0.19$, $SE\ \beta = 0.22$, $\chi^2(1) = 0.77$, $p = .38$), confirming that manipulating both patterns for acoustic variability does not affect generalization. (No main effect was found for the experiment.) Similarly, the interaction between Experiments 3B and 3C was not significant ($\beta = 0.23$, $SE\ \beta = 0.16$, $\chi^2(1) = 0.64$, $p = .43$), confirming that participants generalize to a similar degree for 2:1 and 4:1 contextual variability and exemplar strength advantages.

The mean "yes" rate on generalization items, collapsing across patterns, was higher for Experiment 3B (51.8%) than Experiment 3C (42.1%). This is not surprising, given that Experiment 3C was shorter than its counterpart: the more syllables and repetitions of those syllables to which participants are exposed, the more likely they should be to false alarm on novel syllables. This is consistent with the finding of a significant or marginal main effect of block in most experiments (the only exceptions being Experiments 1B and 3C).

Cross-Experiment Comparison

Our results show a consistently strong role for contextual variability in phonotactic generalization. In Experiment 1, contextual variability was correlated with other factors; in Experiment 2, it was isolated; and in Experiment 3, it was anticorrelated with exemplar strength. In each case, the pattern advantaged for contextual variability was generalized. Three logistic mixed-effects regressions made pairwise comparisons among these experiments to examine any differences in generalization across these three conditions. As shown in Figure 7, there were no significant differences in the rate of generalization across these experiments (interaction of pattern advantage and experiment, 1B vs. 2A: $\beta = 0.31$, $SE\ \beta = 0.21$, $\chi^2(1) = 2.14$, $p = .14$; 1B vs. 3A: $\beta = 0.20$, $SE\ \beta = 0.23$, $\chi^2(1) = 0.79$, $p = .37$; 2A versus 3A: $\beta = -0.11$, $SE\ \beta = 0.20$, $\chi^2(1) = 0.31$ $p = .58$).[1] This suggests that the role of exemplar strength—whether it is correlated, decorrelated, or anticorrelated with contextual variability—is not a crucial factor in generalization.

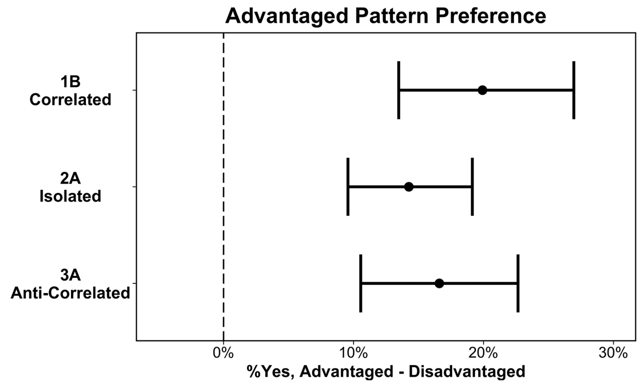

Figure 7. Plot showing results from Experiments 1B (correlated, online), 2A (variability isolated), and 3A (strength and variability anticorrelated, no duration variability). The *x*-axis represents the preference in "yes" response rates for the advantaged pattern: positive values indicate a higher preference for the advantaged pattern. In the case of the anticorrelated experiment, higher values indicate a preference for the variability-advantaged pattern (over the strength-advantaged pattern).

Exemplar Strength and Performance on Familiarization Syllables

Above, learning has been indexed by generalization of patterns to novel linguistic structures. This has long been considered strong evidence that a grammatical rule or constraint has been internalized (e.g., Berko, 1958), and many models of phonotactic learning are tested against word-likeness ratings of nonce words, a type of generalization data (e.g., Albright, 2009; Coleman & Pierrehumbert, 1997; Hayes & Wilson, 2008). That said, the memorization of previously heard words and syllables is a distinct and important measure of learning (Endress & Hauser, 2011), and by this measure exemplar strength has a clear impact: participants are, unsurprisingly, more likely recognize syllables they have heard frequently before. This can be seen in Figure 8, which shows the rate at which each participant correctly recognizes familiarization syllables in Block 2, as a function of how many times they were exposed to those familiarization syllables in Block 1 (that differs across experiment and pattern types). In Block 2, participants have been exposed to all familiarization syllables at least once. If participants have been previously exposed to each syllable of a given pattern *only* once, as in the disadvantaged pattern of Experiments 2B and the advantaged patterns of Experiments 3A and 3B, they correctly recognize those syllables at an average rate of 75.2%. If participants have been previously exposed to those syllables four or more times, they correctly recognize them at an average rate of 98.2%. Participants are clearly tracking exemplar strength, as high exemplar strength leads to recognition of previously seen syllables.

[1] Because of convergence issues, subjects and items were decorrelated for the model comparing Experiments 2A and 3A.

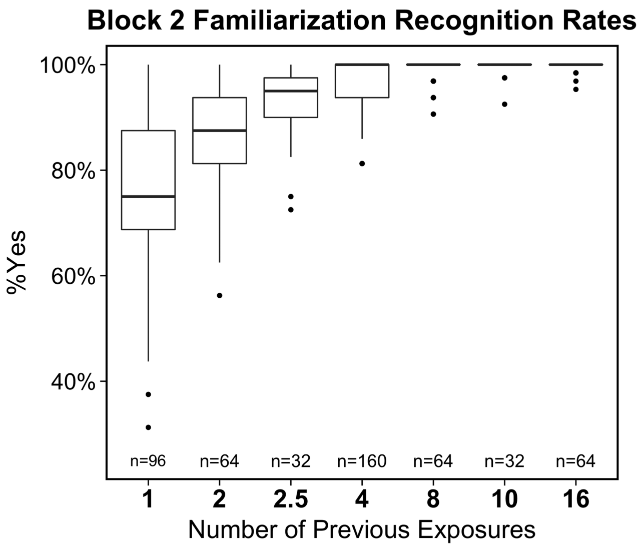

Figure 8. Each participant's correct recognition rates for Block 2 familiarization syllables, by number of previous exposures (i.e., Block 1 exposures). The number of samples per exposure rate is indicated above the *x*-axis.

Participant Exclusion Rates

Ultimately, over half (58.1%) of participants passed the exclusion criteria across all experiments. Inclusion rates were lowest for experiments in which the syllables from one pattern had only a single repetition per block (Experiments 2B, 3A, and 3B). Indeed, among participants who failed the criteria in Experiments 2 and 3, 76.7% of them failed because they did not internalize the familiarization pattern that had fewer repetitions per block. In other words, most of the participants were excluded because they could not have generalized the lower-frequency pattern to novel syllables. In summary, while it is likely that our exclusion criteria resulted in a selection bias toward generally better learners, it is unclear why this would have led to a preference for contextual variability over exemplar strength or vice versa.

General Discussion

Does the learning of gradient phonotactic constraints rely on contextual variability or exemplar strength? Both factors are correlated in natural language, and previous work has failed to adequately decorrelate their contributions. Using an artificial language learning paradigm, this study independently manipulated measures related to exemplar strength and contextual variability (token and type frequency, respectively). Across a series of experiments, we found that contextual variability was a significant factor in learning, while exemplar strength was a much weaker factor. In other words, participants were more likely to extend a phonotactic pattern to novel items if that pattern appeared in a large number of unique syllables; the number of times those syllables were repeated, however, did not appear to affect the rate of generalization.

These results are consistent with previous findings on phonotactics. As reviewed above, studies that have examined the limited number of items where contextual variability and exemplar strength are decorrelated in natural language have argued contextual variability better explains behavior (Buchwald, 2005; Pierrehumbert, 2001). It is also consistent with work from child language acquisition suggesting that greater contextual variability (as measured by larger vocabulary size) attenuates the difficulty of producing relatively infrequent segment sequences (Edwards, Beckman, & Munson, 2004). Our results are also consistent with those of Gerken and Bollt (2008), who found that children generalized natural stress patterns only when the pattern appears with some level of contextual variability.

While our results suggest a strong role for contextual variability regardless of interactions with other factors, Richtsmeier et al. (2011) found that contextual variability is only a significant factor in children's productions when correlated with acoustic variability. This may reflect differences in the mechanisms underlying phonotactic learning in perception and production. For example, artificial language learning paradigms in perception and production have examined the learning of constraints where the distribution of a consonant is conditioned by an adjacent vowel (e.g., if the vowel is /i/, the coda is /f/; if the vowel is /a/, the coda is /n/). In production paradigms, learning such constraints requires more than one session of training to acquire (e.g., Warker & Dell, 2006) but participants can learn such constraints in a single session of perception training (e.g., Steele et al., 2015). The effects of contextual variability may differ in production versus perception in a similar way. Alternatively, the inability of children to generalize based on contextual variability alone in Richtsmeier et al. (2011) could simply be because of the small difference in the number of types between the advantaged and disadvantaged conditions (3 vs. 1).

As noted in the introduction, associative models of phonological learning make strong predictions that exemplar strength should spur generalization (e.g., Goldinger, 1998; Vitevitch & Luce, 1999; Warker & Dell, 2006). Given this, it is striking that contextual variability, rather than exemplar strength, appears to play such a large role in generalization.

To show that our results significantly differ from predictions made by token-based models, we ran simulations with stimuli from two of our experiments using the exemplar model MINERVA 2 (Goldinger, 1998; Hintzman, 1986; details can be found in the Appendix). Using the MINERVA 2 model, we calculated the total activation for each generalization syllable based on previously seen familiarization syllables. Activation of generalization syllables was a function of position-specific segmental overlap with previously seen syllables. Higher total activation (or echo intensity) should result in higher generalization rates (see Hintzman, 1988), as activation reflects the aggregated phonological similarity of the generalization syllable to the set of familiarization syllables.

As discussed earlier, our results from Experiment 2B show the advantaged pattern is not generalized at a higher rate, despite the fact that participants hear advantaged syllables four times per block rather than once per block for disadvantaged syllables. As an example, imagine the advantaged pattern is coda /n/ or /f/, while the disadvantaged pattern is coda /s/ or /b/, and two example generalization syllables are *bef* and *sub*. In such a case, the onset and nucleus of both *bef* and *sub* will each appear an equal number of times (20) in the familiarization syllables for a given block. Coda /f/, however, will appear four times as often as coda /b/ (32 to 8) in a given block, reflecting the difference in exemplar strength between the advantaged and disadvantaged patterns. As can be seen in Figure 9, MINERVA 2 simulations show that generalization syllables reflecting the advantaged pattern, like *bef*, have a mean activation of 25.8, while generalization syllables reflecting the disadvantaged pattern, like *sub*, have a mean activation of 11.6. A linear mixed effects regression was run to determine whether the difference between advantaged and disadvantaged generalization syllables differed, with echo intensity as the dependent measure, and a contrast-coded fixed effect of pattern. Random slopes and intercepts for items and experimental lists were also included.[2] Results indicated the echo intensity was significantly higher for advantaged generalization syllables than disadvantaged ($\beta = 14.1$, $SE\ \beta = 0.42$, $\chi^2(1) = 151.1$, $p < .0001$). In other words, associative models such as MINERVA 2 incorrectly predict increased exemplar strength should lead to a higher degree of generalization.

In Experiment 3C, in which exemplar strength and contextual variability are anticorrelated, participants generalized based on contextual variability, not exemplar strength. Simulations from MINERVA 2 predict the opposite effect from what was observed: mean activation for strength-advantaged generalization syllables was 19.1, while activation for variability-advantaged syllables was only 12.5 (see Figure 10). A linear mixed effects regression, identical to the one described above, revealed this difference was significant ($\beta = -6.6$, $SE\ \beta = 0.40$, $\chi^2(1) = 76.8$, $p < .0001$). Our results, then, appear to be inconsistent with the predictions of associative models.

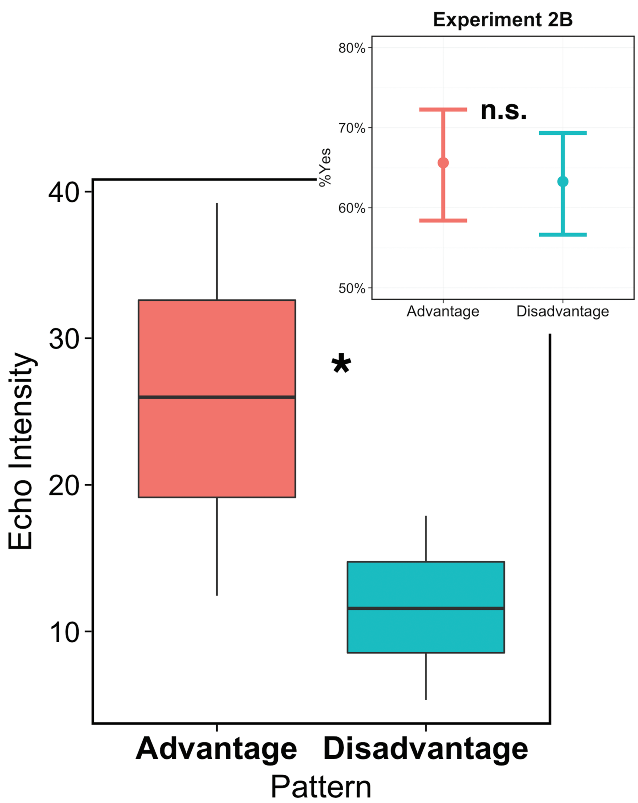

Figure 9. Comparison of echo intensity from MINERVA 2 simulations and behavioral results from Experiment 2B. The top panel reproduces results from Experiment 2B from Figure 5, showing the proportion of "yes" responses on generalization syllables reflecting the advantaged and disadvantaged patterns (error bars reflect bootstrapped 95% confidence interval). The bottom panel is a boxplot, showing the distribution of echo intensity for generalization syllables reflecting the advantaged and disadvantaged patterns. Asterisk indicates significance of pattern in mixed-effects regressions. See the online article for the color version of this figure.

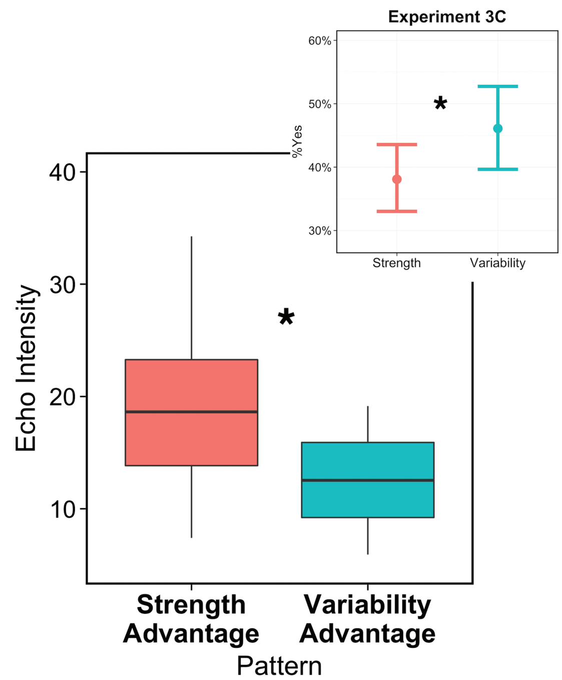

Figure 10. Comparison of echo intensity from MINERVA 2 simulations and behavioral results from Experiment 3C. The top panel reproduces results from Experiment 3C from Figure 6, showing the proportion of "yes" responses on generalization syllables reflecting the advantaged and disadvantaged patterns (error bars reflect bootstrapped 95% confidence interval). The bottom panel is a boxplot, showing the distribution of echo intensity for generalization syllables reflecting the advantaged and disadvantaged patterns. Asterisk indicates significance of pattern in mixed-effects regressions. See the online article for the color version of this figure.

Why is contextual variability so effective in helping learners acquire gradient phonotactic constraints? Such variability may serve to draw the learner's attention away from irrelevant aspects of structure, allowing her to focus on the invariant components that define the phonotactic pattern. For example, in the experiments here, increased variation in onsets and vowels might help learners to focus on those aspects of structure that define the gradient phonotactic constraints: the distribution of coda consonants. This idea draws support from research in nonlinguistic cognitive domains: for example, contextual variability has been shown to be a powerful factor in visual pattern learning, for both adults (e.g., Posner & Keele, 1968) and toddlers (e.g., Quinn & Bhatt, 2010). Within linguistics, contextual variability has been found to enhance pattern learning for a wide variety of phenomena: morphological productivity (Bybee, 1988) as well as morphological pattern learning in children (Nicoladis, Palmer, & Marentette, 2007)

[2] Because of convergence issues, experimental lists and items were decorrelated for the simulations of Experiment 2A.

and adults (Endress & Hauser, 2011); and learning of long-distance syntactic dependencies (Gómez, 2002).

Exemplar strength, on the other hand, has much less support in previous research as a significant factor in pattern learning, despite the general facilitatory effects of frequency in language processing. This may be due, in part, to the exceptional nature of high-strength items; for such items, learners may assume features are idiosyncratic to that item. Consistent with this, exceptions to broad morphological patterns tend to be limited to morphemes with high exemplar strength (as measured by token frequency). Exceptional forms with low token frequency (e.g., the irregular past-tense alternation *weep* - *wept*) are much more likely to be regularized (i.e., changed to *weeped*) than high token frequency forms (e.g., *go* - *went*; Bybee, 1988; see also Bybee, 1995; Bybee & Hopper, 2001; etc.). It should be noted, however, that exemplar strength is crucial for memorization of previously seen items (see previous section for details) and formal models of morphological learning have suggested that performance is best when type and token information are integrated (Goldwater, Griffiths, & Johnson, 2006).

Evidence from computational models of simulated speech communities (Pierrehumbert, 2001) suggests that learners' preference for contextual variability may have a functional motivation. Pierrehumbert argued that for successful communication, phonological patterns must be relatively stable across a speech community. As such, phonological learning must be possible despite the variations in speakers' vocabularies. To examine this issue, she simulated "speakers" with variable vocabularies by randomly sampling from a large English lexicon. Within each simulated speaker, she then examined the distribution of a number of phonological structures. The results revealed that not all "speakers" in the simulated community would acquire the correct patterns if contextual variability was low, even when exemplar strength was extremely high. This suggests that contextual variability may serve an important functional role in the stability of phonological patterns.

We found some evidence that learning increases over time in the experiment, as reflected in significant or marginal interactions between experimental block and pattern in Experiments 2B and 3A. The former finding—in which participants' preference for the strength-advantaged pattern significantly increased over the course of the experiment—may hold implications for theories that posit that exemplar strength plays a crucial role early in the time-course of learning. Pierrehumbert (2003) argues that language learners initiate phonotactic patterns through surface statistics—that is, exemplar strength—and then refine these patterns through "internal feedback from type statistics over the lexicon, once the lexicon is well-developed" (Pierrehumbert, 2003, p. 115). Goldberg, Casenhiser, and Sethuraman (2004), in experimental and corpus investigations, found evidence that a single high-frequency "path-breaking" verb (e.g., *go*, *give*, and *put*) can form a template that aids children learning verb-argument constructions. While these theories suggest that high exemplar strength items may be particularly important for generalization early in learning, our results in fact suggest the opposite: the interaction between block and condition in Experiment 2B reflected that learners' preference for the high-strength pattern *increased* as the experiment wore on.

Conclusions

Previous investigations of phonotactic learning have been unable to dissociate two highly correlated factors present in the input: contextual variability (as indexed by type frequency) and exemplar strength (as indexed by token frequency). In a series of implicit artificial language experiments, this study examined the contributions of both factors. We found that contextual variability is a crucial factor in gradient phonotactic learning, as it consistently affects the rate at which participants generalized familiar phonotactic patterns to novel items. We found no evidence that exemplar strength, on the other hand, significantly affected generalization rates. This is consistent with previous research across multiple domains establishing that variability, by focusing learners' attention on the invariant features of the input, strengthens pattern learning.

References

Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology, 26,* 9–41. http://dx.doi.org/10.1017/S0952675709001705

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Beckman, M. E., & Edwards, J. (2000). Lexical frequency effects on young children's imitative productions. In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology* (Vol. 5, pp. 207–217). Cambridge, United Kingdom: Cambridge University Press.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150–177. http://dx.doi.org/10.1080/00437956.1958.11659661

Bernard, A. (2015). An onset is an onset: Evidence from abstraction of newly-learned phonotactic constraints. *Journal of Memory and Language, 78,* 18–32. http://dx.doi.org/10.1016/j.jml.2014.09.001

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer (Version 5.1.44) [Computer program]. Retrieved from http://www.praat.org/

Buchwald, A. (2005). *Sound structure representation, repair and well-formedness: Grammar in spoken language production* (Unpublished doctoral dissertation). Johns Hopkins University, Baltimore, MD.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. http://dx.doi.org/10.1177/1745691610393980

Bybee, J. L. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology* (pp. 119–141). San Diego, CA: Academic Press.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10,* 425–455. http://dx.doi.org/10.1080/01690969508407111

Bybee, J. L., & Hopper, P. J. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam, the Netherlands: John Benjamins. http://dx.doi.org/10.1075/tsl.45

Chambers, K. E., Onishi, K. H., & Fisher, C. (2010). A vowel is a vowel: Generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology, 36,* 821–828. http://dx.doi.org/10.1037/a0018991

Coleman, J., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. In *Computational phonology: Third meeting of the ACL special interest group in computational phonology* (pp. 49–56). Somerset, NJ: Association for Computational Linguistics.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47,* 421–436. http://dx.doi.org/10.1044/1092-4388(2004/034)

Endress, A. D., & Hauser, M. D. (2011). The influence of type and token frequency on the acquisition of affixation patterns: Implications for language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 77–95. http://dx.doi.org/10.1037/a0020210

Gerken, L., & Bollt, A. (2008). Three Exemplars Allow at Least Some Linguistic Generalizations: Implications for Generalization Mechanisms and Constraints. *Language Learning and Development, 4,* 228–248. http://dx.doi.org/10.1080/15475440802143117

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15,* 289–316. http://dx.doi.org/10.1515/cogl.2004.011

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. http://dx.doi.org/10.1037/0033-295X.105.2.251

Goldrick, M. (2011a). Linking speech errors and generative phonological theory. *Language and Linguistics Compass, 5,* 397–412. http://dx.doi.org/10.1111/j.1749-818X.2011.00282.x

Goldrick, M. (2011b). Using psychological realism to advance phonological theory. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The handbook of phonological theory* (2nd ed., pp. 631–660). Oxford, United Kingdom: Wiley-Blackwell. http://dx.doi.org/10.1002/9781444343069.ch19

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition, 107,* 1155–1164. http://dx.doi.org/10.1016/j.cognition.2007.11.009

Goldwater, S., Griffiths, T., & Johnson, M. (2006). Interpolating between types and tokens by estimating power-law generators. In Y. Weiss, P. B. Schölkopf, & J. C. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 459–466). Cambridge, MA: MIT Press.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436. http://dx.doi.org/10.1111/1467-9280.00476

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry, 39,* 379–440. http://dx.doi.org/10.1162/ling.2008.39.3.379

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551. http://dx.doi.org/10.1037/0033-295X.95.4.528

Hintzman, D. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review, 93,* 411–428. http://dx.doi.org/10.1037/0033-295X.93.4.411

Jusczyk, P. W., & Luce, P. A. (1994). Infants′ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33,* 630–645. http://dx.doi.org/10.1006/jmla.1994.1030

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and language, 37,* 295–311.

Kupin, J. J. (1982). *Tongue twisters as a source of information about speech production*. Bloomington, IN: Indiana University Linguistics Club.

Nicoladis, E., Palmer, A., & Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science, 10,* 237–254. http://dx.doi.org/10.1111/j.1467-7687.2007.00582.x

Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes, 16,* 691–698. http://dx.doi.org/10.1080/01690960143000218

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46,* 115–154. http://dx.doi.org/10.1177/00238309030460020501

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363. http://dx.doi.org/10.1037/h0025953

Quinn, P. C., & Bhatt, R. S. (2010). Learning perceptual organization in infancy: The effect of simultaneous versus sequential variability experience. *Perception, 39,* 795–806. http://dx.doi.org/10.1068/p6639

Richtsmeier, P. T. (2011). Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology, 2,* 157–183. http://dx.doi.org/10.1515/labphon.2011.005

Richtsmeier, P., Gerken, L., & Ohala, D. (2011). Contributions of phonetic token variability and word-type frequency to phonological representations. *Journal of Child Language, 38,* 951–978. http://dx.doi.org/10.1017/S0305000910000371

Seidl, A., Onishi, K. H., & Cristia, A. (2014). Talker variation aids young infants' phonotactic learning. *Language Learning and Development, 5441,* 1–11.

Sommers, M. S., & Barcroft, J. (2007). An integrated account of the effects of acoustic variability in first language and second language: Evidence from amplitude, fundamental frequency, and speaking rate variability. *Applied Psycholinguistics, 28,* 231–249. http://dx.doi.org/10.1017/S0142716407070129

Steele, A., Denby, T., Chan, C., & Goldrick, M. (2015). Learning non-native phonotactic constraints over the web. In The Scottish Consortium for ICPhS 2015. *Proceedings of the XVIIIth International Congress of Phonetic Sciences*. Glasgow, United Kingdom: The University of Glasgow.

Tamaoka, K., & Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavior Research Methods, Instruments & Computers, 36,* 531–547. http://dx.doi.org/10.3758/BF03195600

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (2000). English speakers' sensitivity to phonotactic patterns. *Papers in laboratory phonology V: Acquisition and the lexicon,* 269–282.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 514–529. http://dx.doi.org/10.1037/0278-7393.30.2.514

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9,* 325–329. http://dx.doi.org/10.1111/1467-9280.00064

Vitevitch, M. S., & Luce, P. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408. http://dx.doi.org/10.1006/jmla.1998.2618

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 387–398. http://dx.doi.org/10.1037/0278-7393.32.2.387

Whalen, D. H., Giulivi, S., Nam, H., Levitt, A. G., Hallé, P., & Goldstein, L. M. (2012). Biomechanically preferred consonant-vowel combinations fail to appear in adult spoken corpora. *Language and Speech, 55,* 503–515. http://dx.doi.org/10.1177/0023830911434123

(*Appendix follows*)

Appendix

MINERVA 2 Simulations

Overview of Goldinger's Exemplar Account

In Goldinger's (1998) model, building on Hintzman (1986), recognition memory judgments are related to the strength of the signal a stimulus (*probe*) evokes from memory. Memory consists of a set of *traces*, which are identical in structure to probes. The aggregated similarity of the probe to the set of traces (*echo intensity*) is the strength of the recognition memory signal.

Define a probe p or trace t as N-dimensional vectors. Define the *similarity* of trace t and probe p, S_{tp}, as the dot product of t *and* p, scaled by the vector size $\frac{t \cdot p}{N}$. The *activation* of trace t for probe p A_{tp} is the cube of the similarity S^3_{pt}. The strength of the response from memory, *echo intensity* for probe p with respect to the set of traces $\mathbb{T}$ E^{T}_p, is the sum of activations over all traces: $\Sigma_{i \in \mathbb{T}} A_{ip}$.

In the context of our study, let $\mathbb{P}_\alpha$ and $\mathbb{P}_\delta$ be the sets of probes corresponding to advantaged versus disadvantaged pattern generalization syllables, and $\mathbb{T}_\alpha$ and $\mathbb{T}_\delta$ be the sets of traces corresponding to advantaged versus disadvantaged pattern familiarization syllables. When learning occurs, participants are more likely to say "yes" to generalization syllables from the advantaged versus disadvantaged set. Given that such responses are related to the strength of the response from memory, it must be the case that the echo intensity from items in the advantaged set exceeds that of items in the disadvantaged set:

$$\sum_{i \in \mathbb{P}_\alpha} E_i^{\mathbb{T}} > \sum_{j \in \mathbb{P}_\delta} E_j^{\mathbb{T}}.$$

Prediction: When Exemplar Strength Is Isolated, Learning Should Occur

In Experiment 2B, the sets of generalization and familiarization items have comparable structure across advantaged and disadvantaged syllables—the only distinction is that there are four times as many traces for the advantaged set. There should therefore be a larger echo intensity for the advantaged set, predicting that learning should occur.

Consider one block of training. Divide the set of traces of familiarization syllables from the advantaged pattern into two sets, $\mathbb{T}^F_\alpha$, the first repetition of each syllable, and $\mathbb{T}^R_\alpha$, the remaining traces for advantaged pattern syllables. Note that $\mathbb{T}^F_\alpha$ and $\mathbb{T}_\delta$, the set of traces for disadvantaged pattern syllables, are identical in size. Rewrite the condition for learning in terms of these subsets:

$$\sum_{i \in \mathbb{P}_\alpha} E_i^{\mathbb{T}^F_\alpha} + E_i^{\mathbb{T}^R_\alpha} + E_i^{\mathbb{T}_\delta} > \sum_{j \in \mathbb{P}_\delta} E_j^{\mathbb{T}^F_\alpha} + E_j^{\mathbb{T}^R_\alpha} + E_j^{\mathbb{T}_\delta}.$$

The two sets of generalization items should exhibit comparable echo intensities with respect to the traces of their corresponding familiarization syllables: $\sum_{i \in \mathbb{P}_\alpha} E_i^{\mathbb{T}^F_\alpha} \approx \sum_{j \in \mathbb{P}_\delta} E_j^{\mathbb{T}_\delta}$. Because familiarization syllables in the two sets have comparable distributions of onsets and vowels, differing only in codas, the echo intensities of a probe with respect to traces of familiarization syllables from the nonmatching pattern will also be comparable: $\sum_{i \in \mathbb{P}_\alpha} E_i^{\mathbb{T}_\delta} \approx \sum_{j \in \mathbb{P}_\delta} E_j^{\mathbb{T}^F_\alpha}$. The condition for learning, therefore, reduces to:

$$\sum_{i \in \mathbb{P}_\alpha} E_i^{\mathbb{T}^R_\alpha} > \sum_{j \in \mathbb{P}_\delta} E_j^{\mathbb{T}^R_\alpha}.$$

This condition is clearly met, as echo intensities will necessarily be higher when probes and traces are drawn from matching versus mis-matching sets (as they overlap in coda). Learning is therefore predicted to occur in this context, contrary to what is observed in Experiment 2B.

Simulations of the Exemplar Account

To confirm this analysis, we simulated the actual items used in Experiment 2B, as well as a more complex case (Experiment 3C, where token and type frequency are anticorrelated). For both experiments, we simulated each experimental list that a participant was exposed to. Within each list, we calculated the echo intensity for each probe (i.e., novel generalization syllable). We took the cubed dot product of the probe with each trace (i.e., each previously seen syllable token; note that this included previously seen generalization syllables, as well as familiarization syllables) and summed these to find the echo intensity for each probe. For example, if the probe was *fus*, and one of the traces was *fis*, the dot product of these two syllables is 2/3 (because they overlap on 2 of 3 position-specific segments); cubed, the value was .296. This process was repeated for each trace with the results summed to find the echo intensity of a given probe. Results from simulations of both experiments are reported in the General Discussion.

Received February 29, 2016
Revision received June 20, 2017
Accepted June 23, 2017 ■