

Production and Substantive Bias in Phonological Learning

Youngah Do, Jonathan Havenhill
The University of Hong Kong

1 Introduction

When learning the sound system of a language, learners must induce abstract phonological patterns from heard items that are also suitable for new input (e.g., Pearl & Goldwater, 2016). The induction of abstract phonological patterns, however, does not operate equally for all heard items; experimental evidence suggests that learners are more likely to correctly identify patterns from sounds or groups of sounds exhibiting certain phonological properties (see Moreton & Pater (2012a, 2012b) for a review). With an effort to understand the nature of phonological properties from which learners can better induce phonological patterns, research on phonological pattern induction has primarily tested two hypotheses. First is a structural bias hypothesis suggesting that learners can more easily induce abstract phonological patterns from input that can be classified or defined with fewer phonological features (Moreton & Pater, 2012a). For example, a pattern that classifies [p, t, k] vs. [b, d, g] is easier to induce than a pattern classifying [p, d, k] vs. [b, t, g], because the former classification involves a single phonological feature while the latter involves multiple features (Saffran & Thiessen, 2003). Second is a substantive bias hypothesis proposing that learners better induce phonological patterns that are grounded in ‘phonetic substance’, such as articulatory ease and perceptual salience, than patterns that are not. For example, velar palatalization is more successfully learned when it occurs before a high front vowel than when it occurs before low vowels, because the former process is phonetically grounded in terms of articulation, acoustics, and perception (Wilson, 2006).

Both structural bias and substantive bias have been actively examined in work on phonological learning. However, previous studies have found unequal support for the two biases. The structural bias hypothesis has been consistently supported from a variety of experimental results but studies testing the substantive bias hypothesis have produced mixed results: Some report supporting evidence while others find no difference between learning of substantively grounded and substantively ungrounded patterns. (See the summary of studies in Section 2.) Some studies even report that substantively ungrounded or unnatural patterns are learned better than their grounded or natural counterparts (Albright & Do, 2017; Glewwe et al., 2018). These mixed results raise questions with respect to the strength and the scope of substantive bias in phonological learning. Moreton & Pater (2012b) propose that substantive bias does not affect learning or that its strength is at best weaker than structural bias. Glewwe et al. (2018) attribute the mixed results to an overly broad definition of substantive bias, proposing that substantive bias does affect learning, but that it should be defined only in terms of perceptual factors, excluding articulatory ease. The fact that experimental results have not yet converged toward one explanation suggests that it is imperative to revisit the nature of substantive bias.

The current study explores a novel explanation for the uncertain status of substantive bias in phonological learning. Among the aspects of phonetic substance that may contribute to substantive bias, we focus on articulatory factors. We hypothesize that practice *producing* phonological patterns, a core phonological learning resource alongside perception (Kingston & Diehl, 1994), makes salient to learners the articulatory factors underlying articulatorily (un-)grounded patterns. We predict that the outcome of learning substantively grounded vs. ungrounded patterns will differ when production practice is involved, as opposed to when the patterns are learned entirely through perception. If production is not involved in learning, the

* We would like to thank research assistants May Chan, Suki Chan, Ken Kwong, and Aaron Chik, as well as members of the HKU Language Development Lab. This paper has been improved thanks to comments from the AMP 2020 audience, especially Sara Finley and Eleanor Glewwe.

two patterns will go undifferentiated because the articulatory knowledge constituting the patterns' phonetic grounding has yet to become relevant to learning. Although this hypothesis sounds intuitive, it has several crucial implications for the debate and theories on substantive bias, discussed in Section 2. We explore whether practice with production affects the learning of articulatorily grounded vs. ungrounded patterns, by comparing the outcomes of perception-only vs. perception-with-production learning contexts. Additionally, if production indeed plays a role in learning, its effect should not be restricted to the learning of categorical or absolute phonological patterns. Instead, the effect should also be observed in variable phonological learning. For this reason, we investigate the role of production in variable phonological learning in addition to categorical learning contexts.

2 Production and substantive bias

As introduced above, empirical support for substantive bias has been inconsistent. Considering the mixed results in the field, Moreton & Pater (2012b) propose that substantive bias either does not affect phonological learning or that its strength is at best weaker than that of structural bias. Although it is a plausible argument that can account for the general tendency, we aim to further examine the experimental methodologies used in different studies that may have led to distinct learning outcomes. In other words, before making a wholesale conclusion that substantive bias itself is weak or nonexistent, we examine the testing grounds from which these disparate conclusions have been drawn. The idea is as follows. The effect of substantive bias may not have arisen in some studies, not because the bias itself does not exist, but because the experimental learning conditions did not allow phonetic substance to become relevant to learning. To examine training methodologies employed for different studies, we reviewed previous studies on substantive bias, as shown in Table 1. Our review revealed that there have been two primary methods of training participants, namely training through perception or training that incorporated production along with perception. Notably, the review seems to suggest a link between mode of training and experimental results: Studies with perception-only training have generated mixed results, while studies involving production in training have consistently found evidence for substantive bias, with the exception of Kosa (2010).

On the basis of this potential link, we hypothesize that practice *producing* phonological patterns will facilitate development of articulatory knowledge upon which learners may then base their phonological grammar (Hayes et al., 2004). When articulation is involved in learning, substantively biased learning is likely to be observed. When learning is achieved solely through perceptual input, on the other hand, substantive bias may go unnoticed because the learning mode does not permit learners to recognize the articulatory factors that motivate the sound pattern. If this hypothesis is correct, it may explain why previous studies relying entirely on perception-based training have frequently failed to find evidence for substantive bias. Before testing this hypothesis, we first address the issue of distinguishing articulatory and perceptual grounding of phonological patterns, which are often intertwined.

Although some phonological patterns are motivated both by perceptual salience and articulatory ease, others are more clearly grounded in one domain or the other. Blevins (2008) presents a survey of phonetically natural and unnatural sound patterns, classifying each in terms of whether it is motivated primarily by perceptual factors or by articulatory factors. For instance, she classifies rhinoglottophilia, an observed relationship between aspiration and nasalization, as acoustically and perceptually grounded but with no clear articulatory motivation. Laryngeal and nasal sounds are articulatorily distinct, with the velum and the larynx exerting little, if any, coarticulatory force on one another. However, aspiration and nasalization have similar acoustic effects including increased spectral tilt, increased F1 bandwidth, and increased F1 frequency, such that the two articulations may be perceptually misapprehended for one another (Ohala, 1975). As a result, spontaneous vowel nasalization in the presence of an aspirated consonant (and vice versa) has been observed as a diachronic process in languages including Hindi and Ponapean (Ohala & Busà, 1995; Blevins & Garrett 1993). Similarly, Garellek et al. (2016) show that nasalized vowels tend to be breathier than oral vowels in Yi languages, which they argue is the result of phonetic enhancement.

On the other hand, postnasal voicing, whereby an obstruent following a nasal is preferably voiced rather than voiceless, is a phonetic and phonological tendency driven by aerodynamic-articulatory factors. Raising of the velum in preparation for production of an oral obstruent expands the oral cavity, causing voicing to be maintained into the obstruent closure (Hayes & Stivers, 1996). We are unaware of any reason to suspect that

postnasal voicing is preferable to postnasal devoicing on perceptual grounds. Evidence of postnasal voicing is reflected in a wide variety of languages including Indonesian (Halle & Clements, 1983), Chamorro (Topping, 1969), and Malagasy (Dziwirek, 1989), among others. Numerous languages allow only nasal + voiced obstruent sequences (ND) in the interest of avoiding nasal + voiceless obstruent sequences (NT), including Kikuyu (Clements, 1985) and Oshikwanyama (Steinbergs, 1985). Although some languages, including English, allow both NT and ND sequences, there is no language known to allow NT to the exclusion of ND. This widespread tendency has been formalized as the phonological constraint *NT (Hayes, 1999; Pater, 1996), penalizing nasal + voiceless obstruent sequences. Nevertheless, the phonetically ungrounded pattern of postnasal devoicing does occur in a small number of languages, including Setswana and Sebirwa (Zsiga & Boyer, 2017; Zsiga, 2018), which has been the subject of debate (e.g., Hyman, 2001; Gouskova et al., 2011). As seen in Table 1, learning of grounded postnasal voicing vs. ungrounded postnasal devoicing was tested by Do, Zsiga & Havenhill (2016), who found no evidence for substantively biased learning.

Other alternations, such as velar palatalization, are phonetically grounded in both articulation and perception. In terms of articulation, the competing demands of a velar constriction for [k, g] versus the palatal tongue position for a following [i, j, e] promotes forward movement of the tongue toward the palate (Keating & Lahiri, 1993). This phonetic tendency is clearly observed in the English minimal pair *key* [ki] vs. *coo* [ku]. Initial [k] is typically realized as a front velar before the front vowel in *key*, but as a back velar before the back vowel in *coo*. The process of palatalization is not entirely coarticulatory, however, but involves a perceptual aspect as well. A prominent perceptual cue to a consonant's place of articulation is in the vowel formant transitions into the following vowel, particularly F2. Before a front vowel (which exhibits a high F2), this transition is lost, such that consonants at labial, coronal, and velar places of articulation all exhibit a high F2 characteristic of palatals. Guion (1998) shows that listeners have a tendency to misperceive [k] sequences as [tʃ] due to their perceptual similarity. Thus, a velar consonant in the presence of a high front vowel is both articulatorily and acoustically similar to a palatal consonant. Evidence for substantive bias in the learning of velar palatalization has been found by Wilson (2006), who notably trained participants using a perception-with-production training mode.

Among the studies in Table 1 that have found support for substantive bias using perception-only modes of learning, a large majority have investigated processes of vowel or consonant harmony, which are similarly grounded in both articulatory and perceptual factors (Gallagher, 2010; Kimper, 2017). If learners track a sound's relative perceptibility in various phonological environments, along the lines of the P-map (Steriade, 2001), it is reasonable to expect that they may induce phonological patterns on a perceptual basis even without experience producing that pattern. This seems to be the case for some studies in Table 1 which found support for substantive bias using perception-only modes of learning. Studies that failed to find evidence for substantive bias in perception-only learning tasks have investigated a somewhat wider range of phenomena, some of which have no clear perceptual grounding. For instance, Peperkamp & Dupoux (2007) investigated learning of intervocalic voicing (articulatorily motivated) and found no evidence for substantive bias.

We test the hypothesis regarding the relation between production experience and substantive bias with the case study of postnasal (de)voicing, as previously tested by Do et al. (2016). If substantive bias affects the learning of postnasal (de)voicing regardless of whether production is involved, we predict that the learning of postnasal devoicing will be worse than that of postnasal voicing overall. However, if our hypothesis about the production effect is correct, a difference in learning between postnasal voicing vs. devoicing will be observed through production-based training, but not through perception-based training. If the production effect is robust, we predict that such tendencies will be observed in both categorical and variable pattern learning.

3 Experiments

3.1 Participants 240 adult native speakers of American English were recruited for the experiment through Amazon Mechanical Turk. Recruitment was limited to users with a United States IP address and whose self-reported first language was English. Each participant earned 3 USD upon completion of the experiment. Speakers were excluded from analysis if they did not speak when instructed to during the training phase or if the recording was not successfully captured (20 participants), if they did not correctly respond to the focus questions (1 participant), or if they selected distractor plural forms with an incorrect segment in

more than 30% of trials (8 participants). 28 participants were excluded in total (with one participant satisfying multiple exclusion criteria), resulting in a final dataset with 212 participants.

Mode of training	Study	Phonological pattern	Evidence for substantive bias
With production	Wilson (2006)	Velar palatalization	Yes
	Moreton (2008)	Vowel and consonant harmonies	Yes
	Kosa (2010)	Sibilant harmony	Yes/No
	Moreton (2012)	Vowel and consonant harmonies	Yes
	Stave et al. (2013)	Palatalization	Yes
	White (2014)	Saltatory alternation	Yes
	Strütjen et al. (2018)	Vowel nasalization	Yes
Perception only	Pycha et al. (2003)	Vowel backness (dis)harmony	No
	Wilson (2003)	Nasal (dis)agreement	No
	Peperkamp & Dupoux (2007)	Intervocalic voicing	No
	Zaba (2008)	Vowel and consonant harmonies	No
	Finley & Badecker (2009)	Vowel rounding harmony	No
	Kuo (2009)	Aspiration and place (dis)agreement	No
	Skoruppa & Peperkamp (2011)	Vowel rounding harmony	No
	Do et al. (2016)	Postnasal (de)voicing	No
	Glewwe et al. (2018)	Obstruent voicing contrast	No
	Nevins (2010)	Vowel and consonant harmonies	Yes
	Pons & Toro (2010)	Vowel and consonant harmonies	Yes
	Finley & Badecker (2007)	Vowel rounding harmony	Yes
	Toro, Shukla et al. (2008)	Vowel and consonant harmonies	Yes
	Toro, Nespors et al. (2008)	Vowel and consonant harmonies	Yes
	Zhang & Lai (2010)	Tone sandhi	Yes
	Finley & Badecker (2012)	Vowel height harmony	Yes
	Albright & Do (2017)	Voicing and continuancy alternation	Yes
	Kimper (2016)	Vowel rounding and backness harmony	Yes
	Lin (2016)	Nasal harmony	Yes
	Finley (2017)	Metathesis	Yes

Table 1: Studies testing the substantive bias hypothesis

3.2 Materials Four artificial languages were created following the design of Do et al. (2016). Each language contained 56 (C)VCV training items and 28 (C)VCCV test items, composed of the consonants [p t k b d g f v s z r l j w] and the vowels [i a o u]. The wordlist for each language was balanced such that there was an equal number of voiced and voiceless stops, equal occurrence of each vowel in each vowel slot, and a balanced number of each possible word-initial consonant + vowel pairing. The training items comprised 32

alveolar-initial stems, 16 labial-initial stems, 4 liquid-initial stems, and 4 vowel-initial stems. Velar-initial stems appeared only in the test phase, in order to test whether participants would extend their generalizations to a previously unseen segment. Fricatives and glides appeared only in medial position. The test items were previously unseen words, including 8 labial-, alveolar-, and velar-initial stems (4 voiced and 4 voiceless each), as well as 4 vowel-initial stems. Each test item contained a stop + liquid cluster in medial position, and the vowels were balanced such that for every initial consonant, each vowel appeared once in each vowel slot.

		Naturalness	
		Natural (Voicing)	Unnatural (Devoicing)
Variability	Categorical	/VN + toda/ → [andoda] /VN + tibu/ → [indibu] /VN + dipa/ → [indipa] /VN + dobi/ → [andobi] ...	/VN + toda/ → [antoda] /VN + tibu/ → [intibu] /VN + dipa/ → [intipa] /VN + dobi/ → [antobi] ...
	30% Variable	/VN + toda/ → [andoda] (17/24) /VN + tibu/ → [intibu] (7/24) /VN + dipa/ → [indipa] /VN + dobi/ → [andobi] ...	/VN + toda/ → [antoda] /VN + tibu/ → [intibu] /VN + dipa/ → [intipa] (17/24) /VN + dobi/ → [andobi] (7/24) ...
Filler Stimuli (same for all four languages)		/VN + rulo/ → [irru]lo] /VN + lopa/ → [allopa] /VN + ipu/ → [inipu] /VN + obi/ → [anobi] ...	

Table 2: Breakdown of four artificial languages varying in naturalness (voicing or devoicing) and variability (categorical vs. 70/30 variable), with example stimuli.

During the training phase, participants were instructed to learn the form of a vowel + nasal prefixal plural marker with several allomorphs. First, the vowel of the prefix exhibits height harmony, such that the vowel harmonizes with the height of first vowel of the stem. A stem in which the first vowel is high, such as *ubi*, takes the prefix *in-*, while a stem in which the first vowel is non-high, such as *rasu* or *toba*, takes the prefix *an-*. The prefix also undergoes nasal place assimilation, with the nasal matching the place of articulation of a following stop,¹ as well as total assimilation before liquids, with *l*-initial stems taking the prefix *il-* or *al-* and *r*-initial stems taking the prefix *ir-* or *ar-*. The prefix for vowel-initial stems exhibited height harmony but no change to the nasal.

In addition to these changes in the plural prefix itself, pluralization triggered the target voicing alternation in obstruent-initial stems. This alternation varied between the four languages, both in terms of phonetic naturalness (natural postnasal voicing vs. unnatural postnasal devoicing), as well as whether the alternation applied categorically or variably (applying to only 70% of tokens). In a natural voicing language, the plural of *pabi* was *ambabi*, while in an unnatural devoicing language, the plural of *babi* was *ampabi*. In the variable natural language (VNL), 17 of 24 voiceless-initial stems underwent postnasal voicing (*pabi* → *ambabi*), while 7 of these stems remained faithful to the underlying voicing (e.g., *tabu* → *antabu*). A breakdown of the four languages with example stimuli for each is presented in Table 2.

Stimuli were spoken by an adult female native speaker of American English. All items were recorded in a sound-attenuated booth with a Shure SM58 microphone and Olympus LS-100 solid-state recorder. In order

¹ In practice, nasal place assimilation was observed only for labial-initial stems. Because velar-initial stems were shown only in the test phase, where the plural was presented orthographically (as *in-* or *an-*) but not heard auditorily, it was ambiguous whether the prefix underwent assimilation to [iŋ-] or [aŋ-].

to eliminate any phonetic voicing of postnasal voiceless stops and any devoicing of voiced stops, stimuli were manipulated in Praat (Boersma & Weenink, 2019) to ensure that all voiced stops were fully voiced and that all voiceless stops were fully voiceless. For voiced stops, a portion of the waveform was extracted from the closure of a fully voiced token for each of the voiced stops [b, d, g]. The closure of every voiced stop in the stimulus list was then replaced by the voicing waveform corresponding in place of articulation. The duration of voicing was altered to closely match that of the original stop closure by removing or duplicating individual glottal pulses. For voiceless stops, the stop closures were made fully voiceless by replacing the closure with a duration-matched portion of silence.

3.3 Procedure Participants completed the experiment on a website built using Experigen (Becker & Levine, 2019). Each of the four languages was presented to participants under one of two conditions, perception-only or perception-with-production, for a total of eight experimental settings. After a test to ensure that participants could correctly hear the stimuli, participants were shown examples of how to spell Martian words and then given instructions for the training phase. Participants were told they would learn how to create the plural of Martian words, that the plural marker is added to the beginning of the word, and that “the form of the plural is different for different types of words,” but were not given explicit instruction about which segments or alternation to focus on. Participants in all conditions saw identical instructions, except for whether they were told to speak the plural word out loud or whether they were told to type the plural word on occasion. The experiment then proceeded to training.

Training trials proceeded as follows. A singular word was presented with a randomly chosen picture of a Martian creature from Van de Vijver & Baer-Henney (2011). Participants heard a recording of the singular form in isolation and saw the text “This is one {name}”, with the name of the creature shown orthographically. After clicking to continue, participants heard a recording of the plural form in isolation, along with the text “These are some {PL.name}”, the written plural form, and a picture of the same Martian creature in triplicate. In the perception-only condition, participants usually progressed to the next trial after hearing the plural but were asked to type out the plural as a focus question once per every 7 trials. In these trials, the written plural form was initially hidden, but participants were shown the correct spelling before progressing to the next item. A schematic diagram of these training screens is given in Figure 1.

In the perception-with-production settings, a button appeared with the label “Press to record” after the plural form was presented. After clicking the button, participants were given 3 seconds to speak the plural form out loud, indicated with an on-screen countdown. Production attempts were audio recorded directly in the participant’s web browser using a script written with the Recorder.js JavaScript library (Diamond, 2016). Audio was collected only during the 3-second window, in order to avoid collecting extraneous background noise. At the end of the experiment, participants downloaded their recorded production attempts, which had been concatenated into a single .wav file and uploaded the file to an online drop box. Each recording was checked to verify whether the participant had in fact produced the plural training items. Recording was successful for the majority of the participants (100 out of 120). In some cases, participants ignored the instruction to repeat the plural item and remained silent during the recording phase, even though audio was successfully captured. In other cases, no sound was recorded at all, presumably due to muting or disconnection of the participant’s microphone. In both cases, participants whose production attempts could not be verified were excluded from analysis, as noted above.

During the test phase, participants were asked to choose the correct plural form for 28 previously unseen words. Participants heard a recording of the novel singular item along with a novel picture of a Martian creature and the word shown orthographically. Four choices were then provided, testing both whether participants learned the vowel harmony pattern and whether participants learned the voicing alternation. For instance, four choices were shown for [kugri] in the natural language settings: [ingugri] (correct harmony, correct voicing), [inkugri] (correct harmony, incorrect voicing), [illugri] (correct harmony, incorrect segment), and [allugri] (incorrect harmony, incorrect segment). Participants did not hear recordings of these plural forms, nor were they asked to produce them.

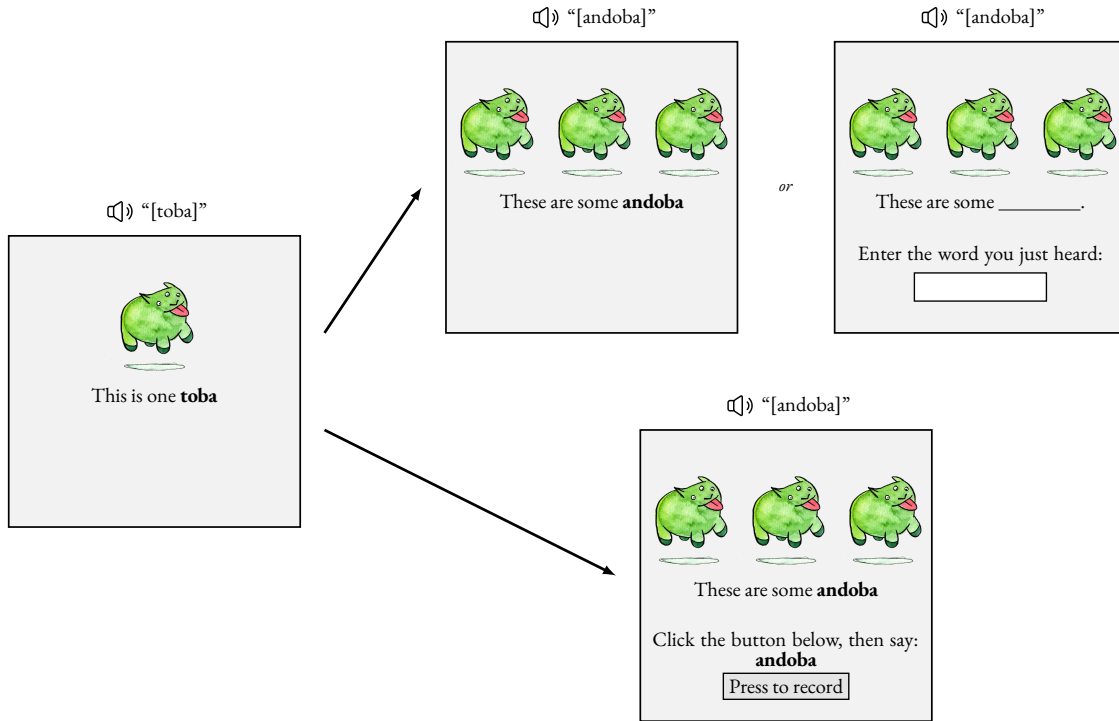


Figure 1: Schematic diagram of training screens for perception-only (top) and perception-with-production (bottom) conditions. Participants were asked to type the plural word (top right) in 8 of 56 trials.

4 Results

Our main prediction concerns the production effect. If practice producing phonological forms during training makes articulatory difficulty relevant to learning, we expect substantively biased learning in the production-based training conditions, but not in the perception-only-based training. In other words, the rate of choosing postnasal (de)voicing will differ between the two conditions depending on production involvement, because learners will recognize articulatory factors that are relevant to postnasal (de)voicing only when they practice producing the forms. Before going into statistical interpretation of the results, we first examine descriptive bar graphs showing postnasal voicing data by production involvement. In Figure 2 and Figure 3, the four languages are shown on the *x*-axis (CNL: categorically natural language; CUL: categorically unnatural language; VNL: variably natural language; VUL: variably unnatural language) and the proportion of postnasal voicing in the participants' responses is shown on the *y*-axis.

We first examine results for the trained segments (labial & coronal), as a measure of explicit learning. In Figure 2, focus on the comparison of dark blue (production) and light green bars (no production) showing the rate of postnasal voicing depending on production involvement. Contrary to our expectation, categorical learning conditions (CNL & CUL) do not seem to show a production effect, as evidenced by similar rates of postnasal voicing between the production vs. no production training conditions. In variable learning conditions (VNL & VUL), however, production involvement increased the rate of choosing postnasal voicing, as shown by the higher rate of voicing in production condition (blue bars) compared to no production condition (green bars). In other words, descriptive data from the explicit learning results show that the production effect was observed only in variable learning conditions.

Next, we examine descriptive results for the untrained segment (velar). This is central to observing how learners generalized the postnasal (de)voicing patterns, because results of the untrained segment allow us to observe patterns of generalization from specific segments toward the broader natural class. The comparison of dark blue bars (production) vs. light green bars (no production) in Figure 3 is relevant. At a descriptive level, the involvement of production appears to facilitate the rate of postnasal voicing production in natural languages (CNL & VNL) but not in unnatural languages (CUL & VUL). In fact, in VUL, the rate of postnasal

voicing was lower when production was involved. In other words, when participants completed production-based training of a language exhibiting more natural patterns (CNL & VNL), they were more likely to generalize the patterns toward a natural alternation pattern, conforming to our prediction regarding the effect of production. If the language itself was unnatural (CUL & VUL), either categorically or variably, production did not increase the likelihood of producing a natural alternation pattern.

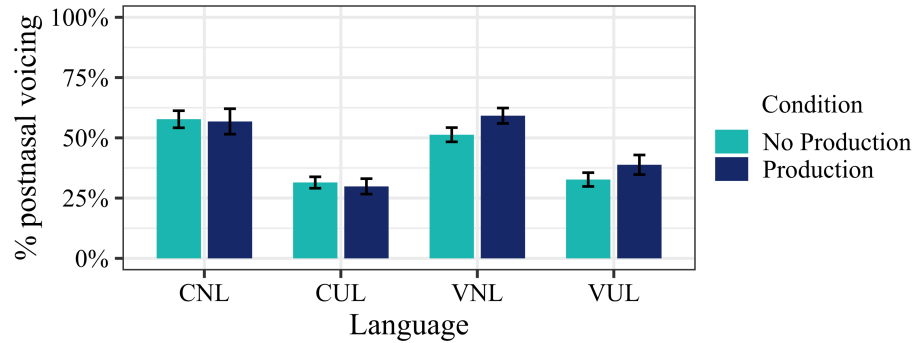


Figure 2: Proportion of postnasal voicing from the explicit learning. Error bars indicate standard error.

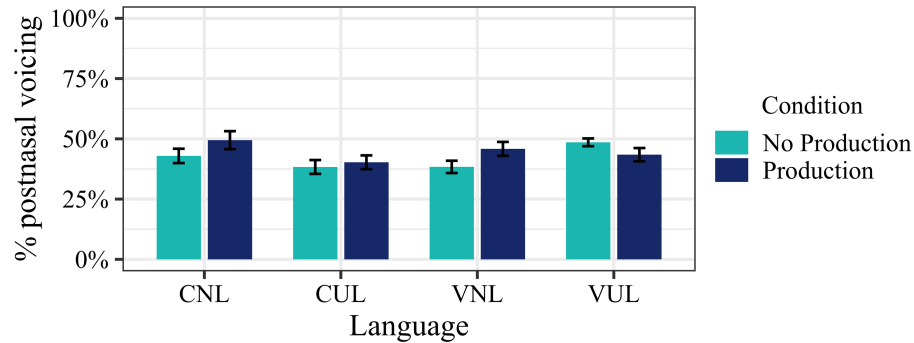


Figure 3: Proportion of postnasal voicing from the generalization test. Error bars indicate standard error.

To test the statistical significance of the generalization test results, we ran two binomial mixed effects logistic regression models using lme4 (Bates et al., 2015), one for the categorical learning condition and another for the variable learning condition. In these models, the dependent variable was the answer choices between the two options, postnasal voicing and postnasal devoicing. We modeled the likelihood of subsequent postnasal voicing responses after learning. The naturalness effect was entered as a binary factor (Natural vs. Unnatural: Baseline), testing whether participants chose postnasal voicing more when trained on natural voicing languages (CNL or VNL) than unnatural devoicing languages (CUL or VUL). The production effect was also entered as a binary factor (Production vs. No production: Baseline), testing whether the choice of postnasal voicing was facilitated specifically when production was involved. Along with the naturalness and production factors, the interaction between the two was incorporated as well, in order to test whether production effect was especially strong when learning natural or unnatural patterns. Random intercepts were included for items and for participants. Model coefficients, along with an assessment of their significance using the Wald-test on z-scores are given in Tables 3 and 4.

	Estimate	Std. Error	z value	p(z)
Intercept	0.06724	0.32486	0.207	0.8360
Natural	0.67449	0.46551	1.449	0.1474
Production	0.43818	0.25262	1.735	0.0828
Natural x Production	0.95162	0.38862	-2.449	0.0143 *

Table 3: Mixed effects model of responses for the generalization test in categorical learning

	Estimate	Std. Error	z value	p(z)
Intercept	-0.65113	0.62881	-1.035	0.3004
Natural	0.09024	0.89089	0.101	0.9193
Production	0.46765	0.19580	2.388	0.0169 *
Natural x Production	0.58343	0.28896	-2.019	0.0435 *

Table 4: Mixed effects model of responses for the generalization test in variable learning

The interpretation of these results is as follows. First, no significant coefficient for naturalness in either model indicates that the naturalness of the language alone did not affect learning performance. This result provides evidence against an overall substantive bias effect. Second, the production effect is only marginally significant in categorical learning conditions (Table 3) suggesting that production did not substantially facilitate the choice of postnasal voicing when the exposed patterns were absolute and deterministic. However, the production factor is significant in variable learning conditions (Table 4): The significantly positive coefficient for Production in Table 4 indicates that production facilitated the choice of postnasal voicing. In other words, a production effect was found only in variable learning conditions. In both categorical and variable learning conditions, the interaction of naturalness and production is significant, and its coefficient is positive. This indicates that the choice of postnasal voicing was facilitated especially when production was included for learning natural languages. The results concur with the descriptive data.

5 General discussion

This paper presents evidence for a production effect in phonological learning. The main finding is that the preference for a natural pattern of alternation (postnasal voicing) was greater when participants generalized patterns after having been trained with production experience. This result supports our prediction that practice producing a pattern makes its articulatory grounding available to learners, which in turn is reflected as substantively biased learning. Contrary to our expectation, however, such production effect was not observed across all learning conditions. First, in the explicit learning condition, the production effect was observed in languages with variable alternation patterns but not with categorical alternation patterns. In other words, when participants were tested on the trained segments after learning a categorical alternation pattern, learning was not modulated by the inclusion of production in training. Second, when participants generalized learned patterns toward the broader natural class (in our study, to velars), a production effect was observed in both categorical and variable learning conditions. Thus, the two conditions in which the production effect was observed were (a) in an explicit test where seen segments showed variable alternation and (b) in a generalization task toward an unseen segment. We interpret these findings to mean that production has an effect on learning when learning involves certain types of *uncertainty*, either due to variability shown in the input or due to unfamiliar segments involved in the test (i.e., they have not seen the exact test segment in the input). When the learning condition and task were relatively transparent, such as when participants were tested on the trained segments after learning their absolute and categorical alternations, no production effect was observed. This study also found that the production effect was found to interact with the pattern's naturalness: the rate of the natural alternation increased when training included production, especially for natural languages. Such tendency was found in both categorical and variable learning conditions. If our argument is correct, this result suggests that production makes the relative articulatory difficulty of postnasal (de)voicing relevant to learning especially when learners are trained on natural languages, such as CNL or VNL. For the first finding, if learners always hear a systematic unnatural alternation, e.g., postnasal devoicing, in a language, they may well consider that only postnasal devoicing occurs in the language even if they have a preference for postnasal voicing. That is, when the pattern is absolute and categorical, unnatural patterns can still be learned; it is only in the learning conditions that involved uncertainty where modifications to the input pattern might occur. The second finding suggests that the proposed production effect arises from settings that are similar to natural language. Languages are likely to be in the shape of our natural artificial languages in that attested alternation patterns are skewed toward naturalness: Languages with unnatural alternations, either categorically or variably, are relatively underrepresented. If so, the fact that a production

effect was observed in a natural language-like experimental setting supports our proposal, suggesting that the proposed production effect is likely to be relevant to natural language acquisition.

We would also note that the current findings have implications for the debate on substantive bias in phonology. First of all, our results are partially in line with the idea of Moreton & Pater (2012a, 2012b) that the strength of substantive bias is weaker than that of structural bias. The current experiments did not include an explicit comparison of the weight of substantive bias vs. structural bias, but our results suggest that phonetic substance is made relevant to phonological learning only when triggered. In the present case, it was the act of production that made articulatory difficulty relevant to learning. In comparison, as shown from previous studies, structural bias has been consistently observed regardless of training method or target learning pattern. Therefore, structural bias is presumably stronger or more systematic than substantive bias and is directly available to learners regardless of whether additional learning mechanisms such as production are involved. Our results, however, challenge the complete exclusion of substantive bias from phonological learning. Substantive bias should be assumed (to some degree) to account for the fact that substantively biased learning was observed in production training conditions. Our finding suggests that an additional factor should be incorporated into the consideration of substantive bias, whereby the bias is made relevant to learning through language experience, such as production in the case of learning articulatorily grounded vs. ungrounded patterns. Our results also counter the proposal to altogether exclude articulatory ease from substantive bias (Glewwe et al., 2018). Glewwe et al. provide experimental evidence showing that an articulatorily difficult pattern (final devoicing) can be learned as readily as an articulatorily easy counterpart (final voicing), from which they argue that substantive bias should be grounded only in perceptual factors. They provide supporting evidence in favor of their proposal by showing equal or even better learning of articulatorily difficult patterns as compared to easy ones (Albright & Do, 2017; Do et al., 2016; Skoruppa & Peperkamp, 2011). This idea differs from the general assumption that substantive bias encompasses both articulatory ease and perceptual salience (Finley & Badecker, 2008; Wilson, 2006). If substantive bias excludes articulation, there is no predicted asymmetry in the learning of postnasal voicing vs. devoicing because the patterns differ primarily by articulatory difficulty. Our results, however, suggest that articulatory components should be assumed in substantive bias, without which we cannot account for the learning asymmetry of postnasal voicing vs. devoicing observed in production conditions. We believe that it is more plausible to assume a substantive bias inclusive of both articulatory and perceptual components, because it better accounts for current results as well as for previous studies showing asymmetric learning of articulatorily grounded patterns (see Section 2). Moreover, as discussed in Section 2, it is not easy to tease apart articulatory factors from perceptual factors for many phonological patterns. Such considerations lead us to propose that substantive bias itself is grounded in both articulatory and perceptual factors, but that the bias must be triggered during learning. In our study, it is the act of production that triggers the relevance of articulatory difficulty for learning. Our study is limited to phonological patterns primarily grounded in articulatory factors, but future studies may consider whether the learning of perceptually motivated patterns is facilitated on the basis of learners' language experience. Additionally, the role of production in learning patterns that are grounded equally on articulation and perception requires further investigation, in order to understand the nature of production experience attested in our study.

References

- Albright, Adam, & Youngah Do. (2017). A substantive bias for perceptually minimal alternations in Artificial Grammar learning. Presentation given at the 14th Old World Conference on Phonology.
- Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Becker, Michael, & Jonathan Levine. (2019). Experigen – an online experiment platform. Available at <http://becker.phonologist.org/experigen>.
- Blevins, Juliette. (2008). Natural and unnatural sound patterns. In Klaas Willems & Ludovic De Cuypere (Eds.), *Naturalness and iconicity in language* (pp. 121–148). Amsterdam: John Benjamins Publishing.
- Blevins, Juliette & Andrew Garrett. (1993). The Evolution of Ponapeic Nasal Substitution. *Oceanic Linguistics*, 32(2), 199–236.
- Boersma, Paul, & David Weenink. (2019). Praat: Doing phonetics by computer (6.0.49).
- Clements, George N. (1985). The problem of transfer in nonlinear phonology. *Cornell Working Papers in Linguistics*, 7.

- Diamond, Matt. (2016). Recorder.js. Available at <https://github.com/mattdiamond/Recorderjs>.
- Do, Youngah, Elizabeth C. Zsiga, & Jonathan Havenhill. (2016). Naturalness and frequency in implicit phonological learning. Presentation given at the 90th Annual Meeting of the Linguistic Society of America (LSA).
- Dziwirek, Katarzyna. (1989). Malagasy Phonology and Morphology. *Linguistic Notes from La Jolla*, 15, 1–30.
- Finley, Sara. (2017). Perceptually based constraints and metathesis: Evidence from artificial grammar. *Proceedings of the 2016 Annual Meeting on Phonology*, 4.
- Finley, Sara, & William Badecker. (2007). Towards a substantively biased theory of learning. *Annual Meeting of the Berkeley Linguistics Society*, 33(1), 142–153.
- Finley, Sara, & William Badecker. (2008). Analytic biases for vowel harmony languages. In Natasha Abner & Jason Bishop (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics* (pp. 168–176). Cascadilla Proceedings Project.
- Finley, Sara, & William Badecker. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, 61(3), 423–437.
- Finley, Sara, & William Badecker. (2012). Learning biases for vowel height harmony. *Journal of Cognitive Science*, 13(3), 287–327.
- Gallagher, Gillian. (2010). Perceptual distinctness and long-distance laryngeal restrictions. *Phonology*, 27(3), 435–480.
- Garellek, Marc, Amanda Ritchart, & Jianjing Kuang. (2016). Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics*, 59, 110–121.
- Glewwe, Eleanor, Jesse Zymet, Jacob Adams, Rachel Jacobson, Anthony Yates, Ann Zeng, & Robert Daland. (2018). Substantive bias and the acquisition of final (de)voicing patterns. Presentation given at the 92nd Annual Meeting of the Linguistic Society of America, Salt Lake City.
- Gouskova, Maria, Elizabeth C. Zsiga, & One Tlale. (2011). Grounded constraints and the consonants of Setswana. *Lingua* 121(15), 2120–2152.
- Guion, Susan Guignard. (1998). The role of perception in the sound change of velar palatalization. *Phonetica*, 55(1–2), 18–52.
- Halle, Morris, & George N. Clements. (1983). *Problem book in phonology: a workbook for introductory courses in linguistics and in modern phonology*. Cambridge, MA: MIT Press.
- Hayes, Bruce. (1999). Phonetically driven phonology: The role of optimality theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan, & Kathleen Wheatley (Eds.), *Functionalism and formalism in linguistics: Volume I: General papers* (pp. 243–285). John Benjamins Publishing.
- Hayes, Bruce, & Tanya Stivers. (1996). The phonetics of postnasal voicing. Ms. University of California, Los Angeles.
- Hayes, Bruce, Robert Kircher, & Donca Steriade. (2004). *Phonetically based phonology*. Cambridge, UK: Cambridge University Press.
- Hyman, Larry M. (2001). The limits of phonetic determinism in phonology: NC revisited. In Elizabeth V. Hume & Keith Johnson (Eds.), *The role of speech perception in phonology* (pp. 141–186). New York: Academic Press.
- Keating, Patricia, & Aditi Lahiri. (1993). Fronted velars, palatalized velars, and palatals. *Phonetica* 50(2), 73–101.
- Kimper, Wendell. (2017). Not crazy after all these years? Perceptual grounding for long-distance vowel harmony. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 19.
- Kimper, Wendell. (2016). Asymmetric generalisation of harmony triggers. *Proceedings of the 2015 Annual Meetings on Phonology*, 3.
- Kingston, John, & Randy L. Diehl. (1994). Phonetic Knowledge. *Language*, 70(3), 419–454.
- Kosa, Loredana Andreea. (2010). Sibilant harmony: Investigating the connection between typology and learnability. *Proceedings of the 2010 Annual Conference of the Canadian Linguistic Association*.
- Kuo, Li-Jen. (2009). The role of natural class features in the acquisition of phonotactic regularities. *Journal of Psycholinguistic Research*, 38(2), 129–150.
- Lin, Yu-Leng. (2016). *Sonority effects and learning bias in nasal harmony*. Doctoral dissertation, University of Toronto, Canada.
- Moreton, Elliott. (2008). Analytic bias and phonological typology. *Phonology*, 25(1), 83–127.
- Moreton, Elliott. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language*, 67(1), 165–183.
- Moreton, Elliott, & Joe Pater. (2012a). Structure and substance in artificial-phonology learning, Part I: Structure. *Language and Linguistics Compass*, 6(11), 686–701.
- Moreton, Elliott, & Joe Pater. (2012b). Structure and substance in artificial-phonology learning, Part II: Substance. *Language and Linguistics Compass*, 6(11), 702–718.
- Nevins, Andrew. (2010). *Locality in vowel harmony*. Cambridge, MA: MIT Press.
- Ohala, John J. 1975. Phonetic explanations for nasal sound patterns. In Charles A. Ferguson, Larry M. Hyman, & John J. Ohala (Eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*. Stanford: Language Universals Project. 289–316.

- Ohala, John J. & Maria Grazia Busà. 1995. Nasal loss before voiceless fricatives: A perceptually-based sound change. [Special issue on *The Phonetic basis of Sound Change*, ed. by Carol A. Fowler] *Rivista di Linguistica* 7. 125-144.
- Pater, Joe. (1996). *Consequences of constraint ranking*. Doctoral dissertation, McGill University.
- Pearl, Lisa, & Sharon Goldwater. (2016). Statistical learning, inductive bias, and bayesian inference in language acquisition. In Jeffrey L. Lidz, William Snyder, & Joe Pater (Eds.), *The Oxford handbook of developmental linguistics* (pp. 664–695). Oxford University Press.
- Peperkamp, Sharon, & Emmanuel Dupoux. (2007). Learning the mapping from surface to underlying representations in an artificial language. In Jennifer Cole & José Ignacio Hualde (Eds.), *Laboratory phonology* (Vol. 9, pp. 315–338). Mouton de Gruyter.
- Pons, Ferran, & Juan M. Toro. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition*, 116(3), 361–367.
- Pycha, Anne, Pawel Nowak, Eurie Shin, & Ryan Shosted. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In Gina Garding & Mimu Tsujimura (Eds.), *WCCFL 22: Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 423–435). Somerville, MA: Cascadilla Press.
- Saffran, Jenny R., & Erik D. Thiessen. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39(3), 484–494.
- Skoruppa, Katrin, & Sharon Peperkamp. (2011). Adaptation to novel accents: Feature-based learning of context-sensitive phonological regularities. *Cognitive Science*, 35(2), 348–366.
- Stave, Matthew, Amy Smolek, & Vsevolod Kapatsinski. (2013). Inductive bias against stem changes as perseveration: Experimental evidence for an articulatory approach to output-output. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35(35), 3454–3459.
- Steinbergs, Aleksandra. (1985). The role of MSCs in Oshikwanyama loan phonology. *Studies in African Linguistics*, 16, 89-101.
- Steriade, Donca. (2001). The phonology of perceptibility effects: The p-map and its consequences for constraint organization. Ms. University of California, Los Angeles.
- Strütjen, Kim, Peter Indefrey, Dinah Baer-Henney, & Ruben van de Vijver. (2018). Perceptual bias in learning a vowel nasalization pattern.
- Topping, Donald M. (1969). A Restatement of Chamorro Phonology. *Anthropological Linguistics*, 11, 62-77.
- Toro, Juan M., Marina Nespov, Jacques Mehler, & Luca L. Bonatti. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, 19(2), 137–144.
- Toro, Juan M., Mohinish Shukla, Marina Nespov, & Ansgar D. Endress. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception & Psychophysics*, 70(8), 1515–1525.
- Van de Vijver, Ruben, & Dinah Baer-Henney. (2011). Acquisition of voicing and vowel alternations in German. In Nick Danis, Kate Mesh, & Hyunsuk Sung (Eds.), *Proceedings of the 35th Annual Boston University Conference on Language Development* (pp. 603–615). Cascadilla Press.
- White, James. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, 130(1), 96–115.
- Wilson, Colin. (2003). Experimental investigation of phonological naturalness. In Gina Garding & Mimu Tsujimura (Eds.), *WCCFL 22: Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 533–546). Cascadilla Press.
- Wilson, Colin. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, 30(5), 945–982.
- Zaba, Aleksandra. (2008). *Relative frequency of patterns and learnability: The case of phonological harmony*. Doctoral dissertation, University of Utah.
- Zhang, Jie, & Yuwen Lai. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(1), 153–201.
- Zsiga, Elizabeth C. (2018). In Favor of [Fortis]: Evidence from Setswana and Sebirwa. *Proceedings of the 2017 Annual Meeting on Phonology*.
- Zsiga, Elizabeth C. & One Tlale Boyer (2017). A natural experiment in learning an unnatural alternation: Sebirwa in contact with Setswana. In J. Kandybowicz and H. Torrence (Eds.), *Africa's Endangered Languages: Documentary and Theoretical Approaches*. Oxford: Oxford University Press.