Complexity Bias and Substantive Bias in Phonotactic Learning

Eleanor Glewwe
eleanorglewwe@ucla.edu
University of California, Los Angeles
LSA 2019 | 4 January 2019

1. Background: Synchronic Learning Biases in Phonological Learning

- To what extent is the phonological typology shaped by synchronic learning biases?
- Synchronic (analytic) bias: Learners biased toward acquiring certain phonological systems over others
  - Complexity bias: bias against formally complex patterns
  - Substantive/naturalness bias: bias against phonetically unnatural patterns
- **Research question**: Does phonetic naturalness bias phonotactic learning?
- **Approach**: Test whether learners reproduce attested and phonetically-motivated phonotactic implicationals in artificial grammar learning (AGL) experiments

1.1 Past Research on Synchronic Biases

- A number of studies have uncovered evidence for **complexity bias**: learners acquire **featurally simpler** phonological patterns better (e.g. Moreton 2008; Hayes et al. 2009; Skoruppa & Peperkamp 2011; Moreton 2012)
- Other studies have found evidence for **substantive bias**: learners prefer to acquire **phonetically natural** patterns and underlearn phonetically unnatural patterns (e.g. Wilson 2006; Becker, Ketrez & Nevins 2011; Becker, Nevins & Levine 2012; Finley 2012; Hayes & White 2013; White 2013)
  - However, some studies that purport to find a naturalness bias could be reinterpreted as having found a complexity bias instead (Becker, Ketrez & Nevins 2011; Hayes & White 2013)
  - Others have a pattern of results that is not fully consistent with a naturalness bias account (Wilson 2006)
- Most of these studies have used an AGL paradigm
- Moreton & Pater's (2012a,b) review of work in this area concluded that there is fairly **robust evidence for complexity bias** but **scant evidence for substantive bias**

1.2 Synchronic Biases in Phonotactic Learning

- Investigations of substantive bias have focused mostly on alternations
- A few studies have tested naturalness bias in phonotactic learning:
  - Support for substantive bias:
    - Underlearning of unnatural phonotactic generalizations supported by the English lexicon (Hayes & White 2013)—however, unnatural generalizations were also more complex
  - Lack of support for substantive bias:
    - Equal learning of natural and unnatural phonotactics (Skoruppa & Peperkamp 2011; Myers & Padgett 2014; Greenwood 2016—casual speech condition)
    - Unnatural phonotactic generalization learned *better* than a natural phonotactic generalization (Greenwood 2016—careful speech condition)
- My approach investigates not just phonotactic restrictions but phonotactic implicationals about the existence of contrasts in different positions
  - Implicitly asking learners to compare existence of contrasts across positions may cause a bias to emerge when simply testing learnability of a specific phonotactic constraint doesn't

2. Experiment 1: Positional Extension of an Obstruent Voicing Contrast

2.1 The Phonotactic Implicational

- If a language contrasts voicing in obstruents word-finally (e.g. /ap/ vs. /ab/), it will contrast voicing in obstruents word-initially (e.g. /pa/ vs. /ba/), but not necessarily vice versa
- Phonetic motivation: cues to obstruent voicing more abundant word-initially than word-finally; in particular, VOT available word-initially but not word-finally (Steriade 1997)
  - Voiced and voiceless obstruents should be more perceptually similar (i.e. harder to distinguish) at the end of a word than at the beginning of a word
  - If voicing contrast exists word-finally, where it is harder to perceive, it should exist word-initially, where it is easier to perceive (T/D# → #T/D)
- Implicational supported by the typology (Steriade 1997, Lombardi 1999)

2.2 Method

- Expose subjects to an obstruent voicing contrast in word-initial or word-final position and test whether they extend the contrast to the other position
- In addition to manipulating position of contrast, manipulated what value voicing “neutralizes” to
- Four training conditions defined on two dimensions: *Trained Contrast Position* and *Trained Neutralization Value*

Table 1: Training Conditions

	#T	#D	T#	D#
#D…{T, D}# (*#T)	✗	✓	✓	✓
#T…{T, D}# (*#D)	✓	✗	✓	✓
#{T, D}…D# (*T#)	✓	✓	✗	✓
*#{T, D}…T# (*D#)*	✓	✓	✓	✗

Bold = Word-final contrast
Italics = Neutralizing-to-T

Table 2: Sample Training Items in the *#{T, D}…T# (*D#)* Condition

#T	#D	T#	D#
pímir **t**ilár **k**awám …	**b**ímir **d**irín **g**awám …	míwi**p** laní**t** nuwá**k** …	

- Properties of items (training and test)
  - $C_1VC_2VC_3$ shape
  - C_1 or C_3 a stop drawn from [p t k b d g]
  - Other two Cs sonorants drawn from [m n l ɹ j w] (no final [j]s or [w]s, no [ji]s or [wu]s)
  - Vs drawn from [i ɑ u]
  - Bilabial, alveolar, and velar stops equally represented
  - Half of items belong to minimal pairs
  - Half iambs, half trochees (stress not correlated with position featuring the voicing contrast or position containing a stop)

- Experiment conducted online using Experigen (Becker & Levine 2013)
- Training Phase
  - Subjects told they would be listening to some words of a new language
  - 2 blocks of the same 36 training items
  - Each training item paired with an image
- Test Phase
  - Subjects heard additional words and had to say whether the word could be a word of the language they had been listening to or not (Yes/No)
  - 1 block of 48 test items: #T, #D, T#, and D# items (same for all conditions)
  - No images
- Three types of test items:
  - *Familiar Conforming*: voicing and position conform to trained pattern, and item heard in training
  - *Novel Conforming*: voicing and position conform to trained pattern, but item not heard in training
  - *Novel Nonconforming*: voicing and position combination not heard in training

Table 3: Sample Test Items for Each Training Condition

	Familiar Conforming	Novel Conforming	Novel Nonconforming
#D…{T, D}# (*#T)	nimá**p**	rínu**p**	**p**írum
#T…{T, D}# (*#D)	nimá**p**	rínu**p**	**b**írum
#{T, D}…D# (*T#)	**k**awám	**k**ámir	múli**k**
*#{T, D}…T# (*D#)*	**k**awám	**k**ámir	múli**g**

2.3 Participants

- Native English speakers recruited through UCLA Psychology Subject Pool
- 149 subjects, divided among 4 conditions (subjects per condition ranged from 33 to 41)

2.4 Predictions

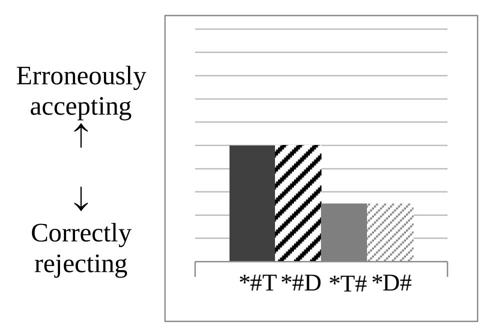
- Subjects' acceptance rates of Novel Nonconforming items (relative to Novel Conforming items) indicate whether they have extended the obstruent voicing contrast to a new position in a given condition

Hypothesis

#1 Substantive bias hypothesis (position—Trained Contrast Position):

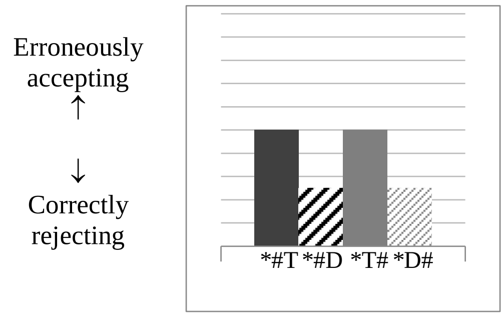
- Recall the phonotactic implicational: T/D# → #T/D, but not vice versa
- Behavior consistent with implicational would be asymmetric extension: subjects exposed to contrast word-finally should extend it to word-initial position **more** than subjects exposed to contrast word-initially extend it to word-final position

Predicted Relative Acceptance Rates of <u>Novel Nonconforming</u> Items by Condition



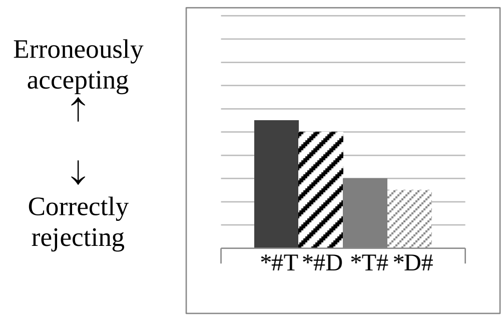
#2 Substantive bias hypothesis
(voicing—Trained Neutralization Value):

- Voiced obstruents more marked than voiceless obstruents → more extension from voiced to voiceless obstruents than from voiceless to voiced obstruents



#3 Substantive bias hypothesis
(position and voicing):

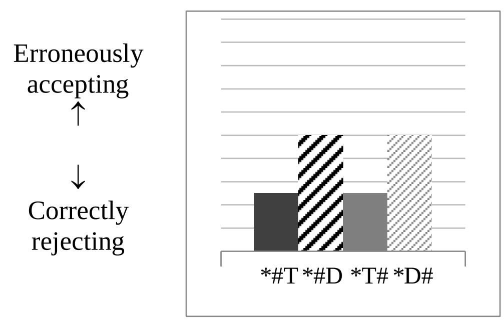
- More extension from word-final to word-initial position AND more extension from voiced to voiceless obstruents



#4 Complexity bias hypothesis:

- Post-hoc, but turned out to provide best account, so presented here for clarity

- Due to presence of sonorant Cs in training items, constraint needed to exclude Novel Nonconforming items in neutralizing-to-T conditions could be more complex than constraint needed in neutralizing-to-D conditions



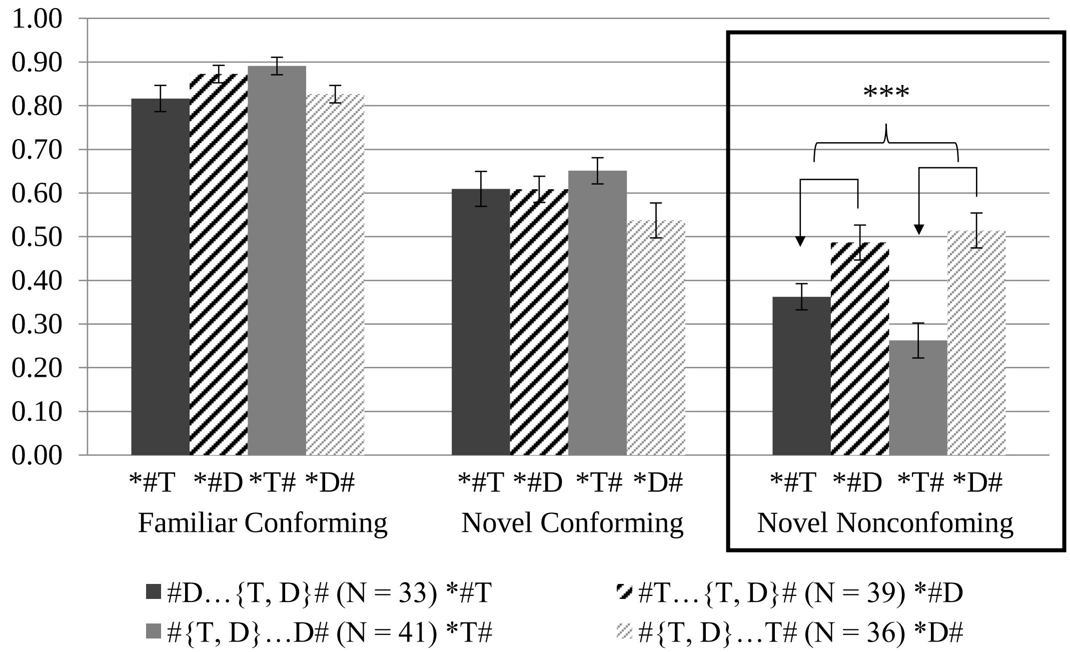
  - ➢ #{T, D}…D# (*T#): kawá**m** ✓ míwi**b** ✓ míwi**p** ✗ → *[–voice]#
  - ➢ *#{T, D}…T# (*D#)*: kawá**m** ✓ míwi**b** ✗ míwi**p** ✓ → *[–son, +voice]#

- Predicts more "extension" from voiceless to voiced obstruents than from voiced to voiceless obstruents (opposite of voicing-related Hypothesis #2 above)

2.5 Results

- Figure 1 shows the acceptance rates of the three types of test items across conditions:

Figure 1: Acceptance Rate of Test Items by Condition



- Acceptance rates of Novel Conforming items:
  - Above chance in all conditions (generalization → learning of trained pattern)
  - Not significantly different across conditions
- Mixed-effects logistic regression fit to Novel Nonconforming items:
  - Dependent variable: response (accept or reject)
  - Fixed effects: Trained Contrast Position and Trained Neutralization Value
  - Random intercepts for subject and item
- Predictions:
  - Position-based substantive bias: Main effect of Trained Contrast Position such that Novel Nonconforming acceptance rates higher in word-final contrast conditions than in word-initial contrast conditions
  - Complexity bias: Main effect of Trained Neutralization Value such that Novel Nonconforming acceptance rates higher in neutralizing-to-T conditions than in neutralizing-to-D conditions

Table 4: Fixed Effects of the Novel Nonconforming Items Regression[1]

	Coefficient	***p***
Intercept	-1.161	<0.001***
Trained Contrast Position = final (vs. initial)	0.197	0.522
Trained Neutralization Value = T (vs. D)	1.063	<0.001***

- Main effect of Trained Contrast Position not significant → no evidence for substantive bias
- Significant main effect of Trained Neutralization Value supports complexity bias

[1] If the interaction of Trained Contrast Position and Trained Neutralization Value is included as a fixed effect in the model, it is not significant ($p = 0.208$).

2.6 Discussion

- The phonotactic implicational (T/D# → #T/D) was not reproduced in this experiment
    - No greater extension of voicing contrast from word-final to word-initial position
    - **Substantive bias hypothesis not supported**
- Instead, subjects trained to "neutralize" to T extended to D more than subjects trained to "neutralize" to D extended to T
    - Opposite of behavior expected based on relative markedness of T vs. D
    - But given (voiced) sonorant Cs, can be explained by a **complexity bias**
    - **#D…{T, D}# (*#T)** and #{T, D}…D# (*T#) subjects can learn *#[–voice] and *[–voice]#
    - But ***#T…{T, D}# (*#D)*** and *#{T, D}…T# (*D#)* subjects must learn more complex *#[–son, +voice] and *[–son, +voice]#
    - Complexity bias account depends on English sonorants having active [+voice] feature
    - However, there are other AGL results consistent with complexity bias story for Exp. 1 (Greenwood 2016, Glewwe et al. 2018)

3. Experiment 2: Modified Voicing Contrast Experiment

3.1 Motivation

- Complexity bias account of Exp. 1: Subjects accepted Novel Nonconforming items less (i.e. learned their language better) when they could use a simpler constraint to master their language
- Which conditions could be learned with simpler constraints depended crucially on features of non-critical Cs of the stimuli, i.e. the voiced sonorants
- To test validity of complexity bias account of Exp. 1, I conducted Exp. 2, which was designed to reverse predictions of complexity bias
- Exp. 2 identical to Exp. 1, except **non-critical Cs converted from voiced sonorants to voiceless fricatives**
- Now it should be easier to learn to reject voiced stops (*D) than to reject voiceless stops (*T)

3.2 Method

- Identical to Exp. 1: Same four training conditions shown in Table 1
- Properties of items (training and test)
    - Identical to Exp. 1, except non-critical Cs (other two Cs in each CVCVC item) drawn from [f θ s ʃ h] ([h] only word-initial)

Table 5: Sample Training Items in the *#{T, D}…T#* Condition

#T	#D	T#	D#
pífis **t**iʃás **k**aʃáf …	**b**ífis **d**isíθ **g**aʃáf …	fýsi**p** ʃaθí**t** θuʃá**k** …	

- Procedure identical to that of Exp. 1

3.3 Participants

- Same population as in Exp. 1

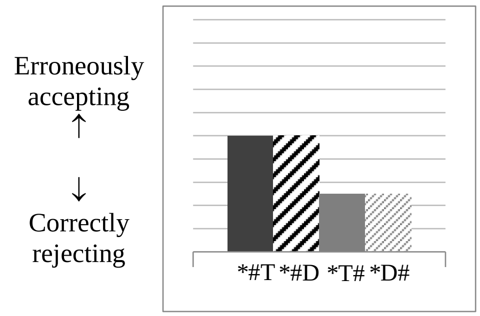
- 144 subjects, divided among 4 conditions (subjects per condition ranged from 35 to 37)

3.4 Predictions

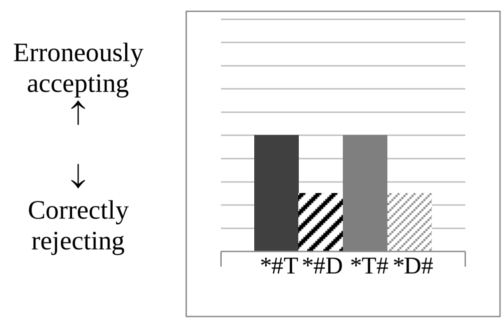
- Like Exp. 1, Exp. 2 still tests for substantive bias, but complexity bias predictions have changed

Hypothesis	**Predicted Relative Acceptance Rates of <u>Novel Nonconforming</u> Items by Condition**

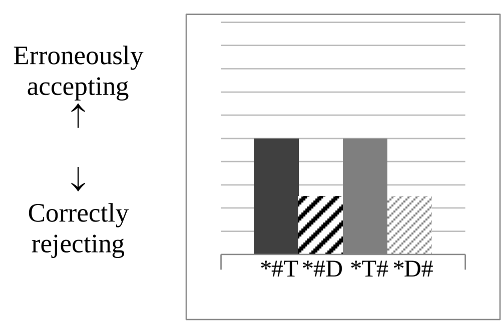
#1 Substantive bias hypothesis (position—Trained Contrast Position):
- Phonotactic implicational: T/D# → #T/D, but not vice versa
- Subjects exposed to contrast word-finally should extend it to word-initial position **more** than subjects exposed to contrast word-initially extend it to word-final position



#2 Substantive bias hypothesis (voicing—Trained Neutralization Value)
- Voiced stops more marked than voiceless stops → more extension from voiced to voiceless stops than from voiceless to voiced stops



#3 Complexity bias hypothesis:
- Due to presence of voiceless fricatives in training items, constraint needed to exclude Novel Nonconforming items in neutralizing-to-D conditions could be more complex than constraint needed in neutralizing-to-T conditions

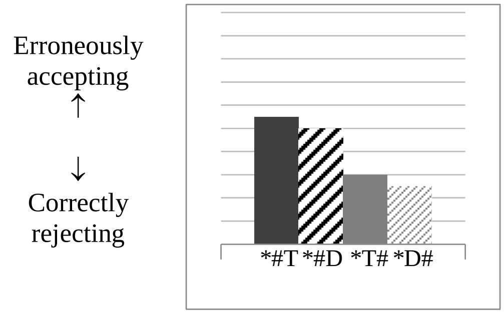


  - #{T, D}…D# (*T#): túsi**f** ✓ físi**b** ✓ físi**p** ✗ → *[–cont, –voice]#
  - *#{T, D}…T# (*D#)*: túsi**f** ✓ físi**b** ✗ físi**p** ✓ → *[+voice]#

- Predicts more "extension" from voiced to voiceless stops than from voiceless to voiced stops
  - **Same prediction** as voicing-related Hypothesis #2 above!
  - Exp. 2 cannot disambiguate between voicing-based substantive bias and complexity bias
  - However, Exp. 1 could, and there was no evidence for voicing-based substantive bias
  - Thus I will interpret this pattern of results in Exp. 2 as supporting complexity bias

#4 Substantive bias hypothesis (position and voicing) OR *Substantive bias hypothesis (position)* ***and*** *Complexity bias hypothesis*

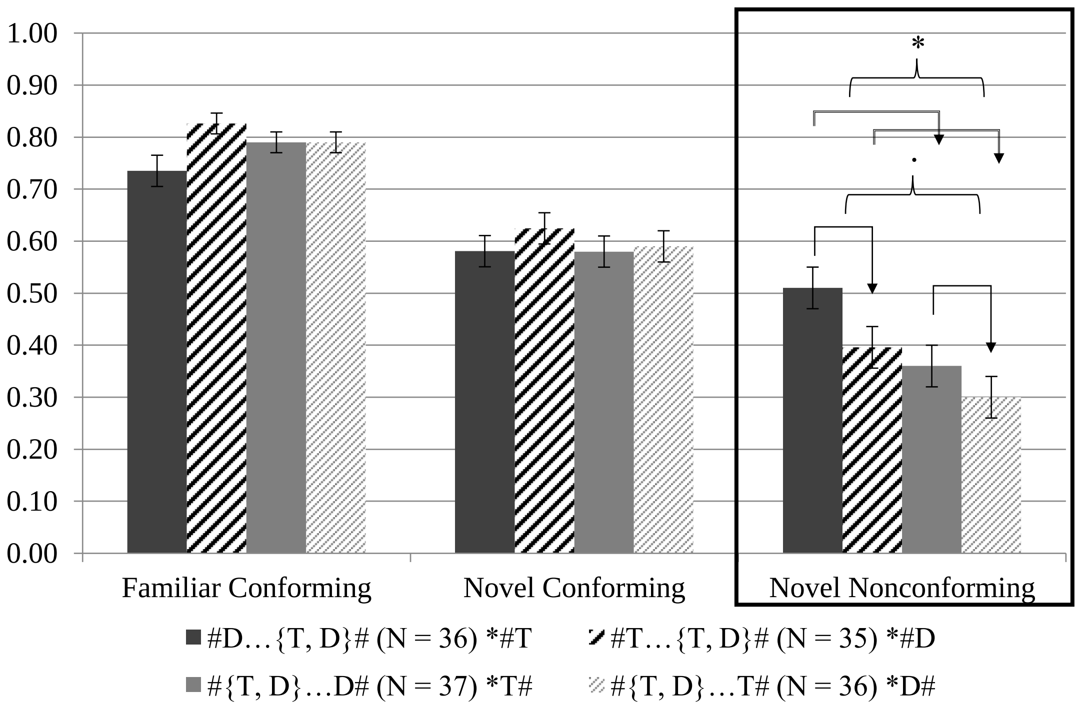
- More extension from word-final to word-initial position AND more extension from voiced to voiceless stops
- I will interpret this pattern of results as supporting position-based substantive bias and complexity bias



3.5 Results

- Figure 2 shows the acceptance rates of the three types of test items across conditions:

Figure 2: Acceptance Rates of Test Items by Condition



- Acceptance rates of Novel Conforming items:
  - Above chance in all conditions (generalization → learning of trained pattern)
  - Not significantly different across conditions
- Mixed-effects logistic regression fit to Novel Nonconforming items:
  - Dependent variable: response (accept or reject)
  - Fixed effects: Trained Contrast Position and Trained Neutralization Value
  - Random intercepts for subject and item
- Predictions:
  - Position-based substantive bias: Main effect of Trained Contrast Position such that Novel Nonconforming acceptance rates higher in word-final contrast conditions than in word-initial contrast conditions

- Complexity bias: Main effect of Trained Neutralization Value such that Novel Nonconforming acceptance rates higher in neutralizing-to-D conditions than in neutralizing-to-T conditions

Table 6: Fixed Effects of the Novel Nonconforming Items Regression[2]

	Coefficient	***p***
Intercept	-1.200	<0.001***
Trained Contrast Position = final (vs. initial)	0.711	0.012*
Trained Neutralization Value = D (vs. T)	0.522	0.065

- Significant main effect of Trained Contrast Position supports substantive bias
- Marginally significant main effect of Trained Neutralization Value supports complexity bias

3.6 Discussion

- Subjects did behave in accordance with the phonotactic implicational (T/D# → #T/D)
  - Greater extension of voicing contrast from word-final to word-initial position
  - **Supports substantive bias**
- Also, subjects trained to "neutralize" to D extended to T more than subjects trained to "neutralize" to T extended to D
  - Higher acceptance of Novel Nonconforming items in conditions where subjects had to learn *[–cont, –voice] than in conditions where subjects could learn *[+voice]
  - **Consistent with complexity bias**
- Flipping voicing of non-critical Cs flipped direction of differences in Novel Nonconforming acceptance rates between neutralizing-to-T and neutralizing-to-D languages
  - Subjects always better at rejecting Novel Nonconforming items featuring stops whose voicing is opposite that of non-critical Cs
  - Non-critical Cs crucially affect AGL performance
  - Subjects infer phonotactic constraints according to experiment-internal distribution of sounds, opting for simplest constraint with which they can master pattern

4. General Discussion

- Two experiments tested the phonotactic implicational whereby a word-final obstruent voicing contrast entails a word-initial contrast, but not vice versa
  - **Exp. 1**: **Natural languages** with a stop voicing contrast only word-initially **not reliably learned better than unnatural languages** with a stop voicing contrast only word-finally
  - Exp. 2: **Natural languages** with a stop voicing contrast only word-initially **learned better than unnatural languages** with a stop voicing contrast only word-finally
  - **Mixed evidence for substantive bias**
- What is the outlook for substantive bias?
  - Results of Exps. 1 and 2 in line with Moreton & Pater (2012a,b): substantive bias elusive
  - Typological asymmetries rooted in phonetic naturalness must still be accounted for
  - Substantive bias's major competitor in this debate: the diachronic explanation/channel bias (Blevins 2004, Moreton 2008)
    - Phonetic factors (e.g. articulatory difficulty, perceptibility) drove imperfect transmission of languages over time, yielding the present typology

[2] If the interaction of Trained Contrast Position and Trained Neutralization Value is included as a fixed effect in the model, it is not significant ($p = 0.699$).

  - Substantive bias may be subtle/hard to detect, yet notable AGL results exist (Finley 2012, White 2013)
  - Perhaps a difference between articulatory substantive bias and perceptual substantive bias? (Glewwe et al. 2018)
    - In Exps. 1 and 2, position-based substantive bias perceptual while voicing-based substantive bias articulatory, and only position-based substantive bias found any support
    - Naturalness arguments in Finley 2012 and White 2013 rooted in perception
    - Glewwe et al. 2018: naturalness argument articulatory (final devoicing vs. final voicing) and no evidence for substantive bias
- Both experiments also provided **evidence for complexity bias**
  - Subjects accepted Novel Nonconforming items less (i.e. demonstrated **superior learning**) **when pattern could be mastered with a featurally simpler phonotactic constraint**
  - Effect more consistent and robust across two experiments than substantive bias effect
  - Held true whether simpler conditions were neutralizing-to-D conditions (Exp. 1—sonorant non-critical Cs) or neutralizing-to-T conditions (Exp. 2—voiceless fricative non-critical Cs)
  - Exps. 1 and 2 highlight decisive role non-critical segments/phone inventory of an artificial language can play
  - Interpretations of AGL results must take potential role of non-critical sounds into account

Acknowledgments

Many thanks to Kie Zuraw, Megha Sundara, Robert Daland, and Bruce Hayes for their guidance and advice. Thanks also to Adam Chong, Elliot Moreton, and audiences at CLS 53, the Southern California Meeting on Phonology 2017, AMP 5, AMP 6, the Berkeley Phorum, and the UCLA Phonology Seminar.

References

Becker, M., Ketrez, N. & Nevins, A. (2011). The Surfeit of the Stimulus: Analytical Biases Filter Lexical Statistics in Turkish Laryngeal Alternations. *Language*, *87*(1), 84–125.

Becker, M. & Levine, J. (2013). Experigen – an online experiment platform. http://becker.phonologist.org/experigen.

Becker, M., Nevins, A., & Levine, J. (2012). Asymmetries in Generalizing Alternations To and From Initial Syllables. *Language*, *88*(2), 231–268.

Blevins, J. (2004). *Evolutionary phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.

Finley, S. (2012). Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and Cognitive Processes*, *27*(10), 1550–1562.

Glewwe, E., Zymet, J., Adams, J., Jacobson, R., Yates, A., Zeng, A., & Daland, R. (2018). Substantive Bias and Word-Final Voiced Obstruents: An Artificial Grammar Learning Study. Talk at the *92nd Annual Meeting of the Linguistic Society of America*. Salt Lake City.

Greenwood, A. (2016). *An Experimental Investigation of Phonetic Naturalness*. Ph.D. dissertation. University of California, Santa Cruz.

Hayes, B., Siptár, P., Zuraw, K., & Londe, Z. (2009). Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language*, *85*(4), 822–863.

Hayes, B. & White, J. (2013). Phonological Naturalness and Phonotactic Learning. *Linguistic Inquiry*, *44*(1), 45–75.

Jun, J. (1995). *Perceptual and Articulatory Factors in Place Assimilation: An Optimality Theoretic Approach*. Ph.D. dissertation. University of California, Los Angeles.

Lombardi, L. (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory*, *17*(2), 267–302.

Lombardi, L. (2001). Why Place and Voice Are Different: Constraint-Specific Alternations in Optimality Theory. In L. Lombardi (Ed.), *Segmental Phonology in Optimality Theory: Constraints and Representations* (pp. 13–45). Cambridge: Cambridge University Press.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*(1), 83–127.

Moreton, E. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language*, *67*(1), 165–183.

Moreton, E. & Pater, J. (2012a). Structure and Substance in Artificial-phonology Learning, Part I: Structure. *Language and Linguistics Compass*, *6*(11), 686–701.

Moreton, E. & Pater, J. (2012b). Structure and Substance in Artificial-phonology Learning, Part II: Substance. *Language and Linguistics Compass*, *6*(11), 702–718.

Myers, S. & Padgett, J. (2014). Domain generalisation in artificial language learning. *Phonology*, *31*(3), 399–433.

Skoruppa, K. & Peperkamp, S. (2011). Adaptation to Novel Accents: Feature-Based Learning of Context-Sensitive Phonological Regularities. *Cognitive Science*, *35*, 348–366.

Steriade, D. (1994). Positional Neutralization and the Expression of Contrast. Ms. University of California, Los Angeles.

Steriade, D. (1997). Phonetics in phonology: The case of laryngeal neutralization. Ms. University of California, Los Angeles.

Steriade, D. (2001a). Directional asymmetries in place assimilation: A perceptual account. In E. Hume & K. Johnson (Eds.), *The Role of Speech Perception in Phonology* (pp. 219–250). San Diego: Academic Press.

White, J. (2013). *Bias in phonological learning: Evidence from saltation*. Ph.D. dissertation. University of California, Los Angeles.

Wilson, C. (2006). Learning phonology with a substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945–982.