Title
Bias in Phonotactic Learning: Experimental Studies of Phonotactic Implicationals

Author
Glewwe, Eleanor

Publication Date
2019

UNIVERSITY OF CALIFORNIA

Los Angeles

Bias in Phonotactic Learning:

Experimental Studies of Phonotactic Implicationals

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy
in Linguistics

by

Eleanor Rachel Glewwe

2019

ABSTRACT OF THE DISSERTATION

Bias in Phonotactic Learning:

Experimental Studies of Phonotactic Implicationals

by

Eleanor Rachel Glewwe

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2019

Professor Kie Ross Zuraw, Chair

An ongoing debate in phonology concerns the extent to which the phonological typology is shaped by synchronic learning biases. The two best-studied types of synchronic bias are complexity bias, a bias against formally complex patterns, and substantive bias, a bias against phonetically unnatural patterns. While most previous work has focused on bias in the learning of phonological alternations, this dissertation tests for substantive bias and complexity bias in phonotactic learning.

Four artificial grammar learning (AGL) experiments tested whether learners reproduce phonetically-motivated phonotactic implicationals from the typology. The implicationals concern the distribution of place of articulation and voicing contrasts in stops across positions. If a language has place or voicing contrasts in stops word-finally, it also has those contrasts word-initially, but if a language has such contrasts word-initially, it does not necessarily have them word-finally. These implicationals are phonetically motivated: stop place of articulation and voicing are less perceptible word-finally than word-initially. If place or voicing contrasts exist in a position where they are hard to perceive, they should also exist in positions where they are easier to perceive. My experiments exposed participants to place or voicing contrasts in word-initial or word-final position and then tested whether they extended the contrast(s) to the other word-edge position. Perception-based substantive bias predicts greater extension from word-final to word-initial position than vice versa. This prediction was not borne out in the place experiments but was borne out in one voicing experiment. The voicing experiments thus provide partial support for substantive bias. Due to the phonemic inventories of the artificial languages, the voicing experiments could also test for complexity bias. Effects of complexity bias emerged in both experiments.

A fifth AGL experiment tested the relative learnability of final voicing alternations. The experiment failed to find support for articulation-based substantive bias: final devoicing, which increases articulatory ease, was not learned better than final voicing. The results did provide additional support for complexity bias. Based on the results of this dissertation's experiments and a review of the literature, I argue for distinguishing between perceptually-rooted and articulatorily-rooted substantive bias and claim that only perceptual naturalness biases phonological learning.

The dissertation of Eleanor Rachel Glewwe is approved.

Bruce P. Hayes

Donka Minkova Stockwell

Megha Sundara

Kie Ross Zuraw, Committee Chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Many people deserve thanks for making my six years at UCLA a stimulating, fulfilling, and happy time. I am deeply grateful to my advisor, Kie Zuraw, who has gently guided me since my first year and thanks to whose mentorship I have gone from linguistics graduate student to grown-up linguist. I have benefitted enormously from Kie's awe-inspiring teaching, tremendous intellect, sensible advice, and generosity with her time. Kie is unfailingly kind and encouraging, and she was supportive when I passed through the Slough of Despond. When it comes to teaching and advising, Kie is my model. For some strange reason, people sometimes think we look alike; it is an honor to be her pseudo-doppelgänger.

I have also benefitted greatly from the stellar teaching and advising of Bruce Hayes. His deep insights into phonological theory always helped me move my work forward. I appreciate his intellectual deftness and thoroughly Hayesian way of expressing things. I especially thank him for serving as my de facto advisor in my last year.

Thanks to Megha Sundara for sharing her experimental expertise and advice on this dissertation and other projects. Experiment 4 was her idea. I also learned a great deal from Megha in the many courses I took with her. Thanks to Robert Daland, whose excellent Experimental Phonology class was the field from which the seed of this dissertation grew. He was a long-time member of my committee and provided encouragement in a time of discouragement. Thanks to Donka Minkova, who graciously agreed to join my committee in the final months and who taught me about the history of English over chocolate in her office.

Thanks to Jesse Zymet, Robert Daland, Jacob Adams, Rachel Jacobson, Tony Yates, and Ann Zeng for letting me incorporate our group experiment into this dissertation.

Thanks to Pat Keating for taking me to the Thursday shape note singing before I'd even attended my first class at UCLA and for the Children's Book World connection. Thanks to Hannah Sarvasy for being a dazzling inspiration, for making my first academic conference a camping trip, and for staying in touch over the years.

Thanks to my linguistics professors at Swarthmore, especially Vera Lee-Schoenfeld, David Harrison, Donna Jo Napoli, Nathan Sanders, Daniel Altshuler, and Ted Fernald, for showing me that linguistics really was as cool as I had hoped.

Thanks to my parents for laying the foundation for my education and supporting me in every path I have chosen. I also thank them, as well as my brother and my extended family, for their encouragement and confidence in me.

Various communities have made my time in Los Angeles happier. Thanks to Vrinda Chidambaram, the Johnny Appleseed of Georgian choirs, for founding Datvebis Gundi, the UCLA Kartvelian Chorus, and introducing me to the joys of Georgian polyphony. Thanks also to the many Georgian chorus members from over the years who helped fill my Friday evenings with song. Thanks to the local shape note singers, who welcomed me into their midst and without whom I would not have encountered the Korean-American Friendship Bell or the Topanga Banjo-Fiddle Contest. Thanks to the denizens of the UCLA Phonetics Lab for making the lab a nice place to come to work and for much lively conversation over lunch. And thanks to everyone in the department for not once making me feel like I was squandering my time by writing novels and for coming to my book release parties.

Thanks to my friends—from college, high school, and even earlier—who have shared the Ph.D. journey with me, whether ahead of me or behind me on the road: Dustin Anderson, Andrew Cheng, Natalie Heer, Leland Kusmer, Hana Maruyama, Elissa Wong, and Rebekah

Yang. Thanks especially to Dustin for Los Angeles meet-ups and to Leland for an enduring correspondence and two All-Cals (I'm sorry I still haven't made it to WMSHC).

Thanks to my fellow grad students at UCLA, especially Dustin Bowers, Margit Bowler and John Gluckman (for ACAL 48), Canaan Breiss, Adam Chong, Meaghan Fowlie (who along with Chase enticed me to UCLA with singing), Laura Kalin (for coming to the first-years' birthday party), Michael Lefkowitz, Nicoletta Loccioni, Iara Mantenuto, Laura McPherson, Sözen Özkan and Deborah Wong (my birthday buddies), Christine Prechtel, Brice Roberts, Adam Royer, Jeremy Steffman, Beth Sturman, Adeline Tan, Yu Tanaka, Jos Tellings (for the concert at the Mojica Hacienda), Meng Yang, and Jesse Zymet. Their conversation and camaraderie made my time at UCLA richer. I have no doubt forgotten some people, so I hereby thank anyone who ever brightened my day.

Thanks to Chase O'Flynn (Kaccāyana) for Camp Kiya, deep conversations, jam sessions, wisdom and compassion, the hammered dulcimer, inviting me to Dhammadharini, and generally being a treasured friend at UCLA and beyond.

Thanks to Isabelle Lin for endless encouragement and steadfast support, and for believing in me when I almost lost faith. Thank you for all the adventures and conversations, the zines and the short stories, the plays and the concerts, the galleries and the bookstores, the egg tarts and the countless cups of Harney & Sons Paris. Thank you for sharing Æncre, the best godcat in the world. Thank you for Cedar City and Minneapolis, Lisbon and Paris. Ton amitié est la meilleure chose que UCLA m'a apportée. Merci pour tout.

This dissertation was supported by a UCLA Dissertation Year Fellowship.

VITA

2012	B.A. with High Honors, Linguistics and Languages (French and Mandarin) Swarthmore College
2014	Graduate Summer Research Mentorship University of California, Los Angeles
2015	M.A., Linguistics University of California, Los Angeles
2015-2016	Graduate Research Mentorship University of California, Los Angeles
2017-2018	Dissertation Year Fellowship University of California, Los Angeles
2014-2019	Teaching Assistant, Associate, Fellow/Instructor University of California, Los Angeles

PUBLICATIONS

Glewwe, E. (in press). Efik Nominal Tonal Alternations as Phrasal Morphology. In S. Lotven, S. Bongiovanni, P. Weirich, R. Botne, & S. G. Obeng (Eds.), *African linguistics across the disciplines: Selected papers from the 48th Annual Conference on African Linguistics*. Berlin: Language Science Press.

Glewwe, E. (2019). Substantive bias in phonotactic learning: Positional extension of an obstruent voicing contrast. In *Proceedings of CLS 53 (2017)*, 103–116. Chicago, IL: Chicago Linguistic Society.

Glewwe, E. (2017). An OT Analysis of Efik Contrastive Verbal Reduplication. In K. Jesney, C. O'Hara, C. Smith, & R. Walker (Eds.), *Supplemental Proceedings of the 2016 Annual Meeting on Phonology*. Washington, D.C.: Linguistic Society of America.

Glewwe, E. & Aly, A. (2016). Coronal palatalization in Logoori. In D. L. Payne, S. Pachiarotti, & M. Bosire (Eds.), *Diversity in African languages: Selected papers from the 46th Annual Conference on African Linguistics*, 165–183. Berlin: Language Science Press.

Glewwe, E. (2016). Tonal adaptation in English loanwords in Mandarin: A corpus study. In *Proceedings of CLS 51 (2015)*, 197–211. Chicago, IL: Chicago Linguistic Society.

CHAPTER 1: INTRODUCTION

A current subject of debate in phonological theory is the extent to which the phonological typology is shaped by synchronic learning biases. Synchronic bias, also called analytic bias (Moreton 2008), refers to the notion that learners are biased toward acquiring certain phonological systems over others. Two main types of synchronic bias have been discussed and tested in the literature: complexity bias, a bias against formally complex patterns, and substantive bias (also called naturalness bias), a bias against phonetically unnatural patterns. This dissertation investigates the roles of substantive bias and complexity bias in phonotactic learning. I approach the question of whether phonetic naturalness biases phonotactic learning by testing whether learners reproduce attested and phonetically motivated phonotactic implicationals in artificial grammar learning (AGL) experiments. One implicational involves the distribution of major place contrasts across word positions while the other involves the distribution of an obstruent voicing contrast across word positions. An additional experiment testing the learning of phonological alternations compares the learnability of word-final obstruent voicing and devoicing and complements the phonotactic learning experiments on the obstruent voicing contrast. Taken together, the experimental results provide mixed evidence for substantive bias but more robust evidence for complexity bias. They also suggest that substantive bias rooted in perceptual naturalness holds more promise than substantive bias rooted in articulatory naturalness.

1.1 Testing for Synchronic Learning Biases in Phonology

1.1.1 Methodologies

Studies testing for synchronic biases in phonological learning take one of two broad approaches. The first approach, variously called L1 (first-language) or natural language studies (Prickett 2015, 2018), generalization studies (Baer-Henney 2015), or the nonce-word query paradigm (White 2013), looks for bias by probing what learners have internalized from the lexicons of their own native languages. Experiments present participants with nonce words of their native language (that is, unattested words that conform to the language's phonotactics to some desired degree) and ask them to judge their well-formedness (Scholes 1966) or apply a morphological rule (e.g. plural formation) to them that requires or gives participants the opportunity to apply a phonological alternation. Well-formedness judgment tasks are sometimes called blick tests (Chomsky & Halle 1965, Chong 2017) while tasks that ask participants to apply a morphological rule are a type of wug test (Berko 1958). (Note that the terms blick test and wug test are not used exclusively for native language experiments but have also been applied to tasks following the training phase in artificial grammar learning studies, as in Chong 2017.)

Under these two categories, specific possible tasks include rating words' goodness (Hayes et al. 2009, Hayes & White 2013, Prickett 2015, 2018), choosing between two proposed forms (two-alternative forced choice) (Becker, Ketrez & Nevins 2011), expressing a preference between two proposed forms on a scale (Becker, Clemens, & Nevins 2017), written production (Becker, Nevins & Levine 2012), and spoken production (van de Vijver & Baer-Henney 2014). Using such tasks, these studies measure native speakers' internalization of their language's phonological patterns and compare them to these patterns' prevalence in the language input. Mismatches between how well a pattern is learned and how much evidence supports it in the input can point to learning biases. For instance, if a simple and a complex pattern are equally well-attested but participants' behavior shows they have learned the simple one better, this suggests that complexity constrains phonological learning.

Past native language studies investigating synchronic bias include blick tests on English (Hayes & White 2013, Prickett 2015, 2018) and wug tests on Hungarian (Hayes et al. 2009), Turkish (Becker, Ketrez, & Nevins 2011), English (Becker, Nevins, & Levine 2012), German (van de Vijver & Baer-Henney 2014), French (Becker, Clemens, & Nevins 2017), and Portuguese (Becker, Clemens, & Nevins 2017).

The second and more common approach to testing for synchronic bias is artificial grammar learning (AGL), in which participants are exposed to a phonological pattern absent from their native language and then tested on their learning of the pattern. In such studies, experimenters can carefully control the artificial languages taught to participants so that they differ in exactly and only the property of interest, such as featural complexity or phonetic grounding. Since participants are usually assumed to lack any native-language experience that would influence their learning of the experimental patterns, a difference in learning suggests an effect of some type of bias. Of course, the potential influence of participants' native language on their performance in AGL experiments cannot be ignored, so some studies take care to show that native language factors such as lexical statistics could not furnish an adequate alternative explanation for their findings (e.g. Moreton 2008, Finley & Badecker 2012, Myers & Padgett 2014) or even treat native language factors as an experimental variable (Baer-Henney, Kügler, & van de Vijver 2015). Like natural language studies, AGL experiments can involve phonotactics (e.g. Myers & Padgett 2014) or alternations (e.g. Pycha et al. 2003). They can also test learning with a variety of tasks, including two-alternative forced choice (e.g. Moreton 2008, Finley 2008, White 2013), deciding whether an item belongs to the training language or not (e.g. Pycha et al. 2003, Myers & Padgett 2014), written production (e.g. Becker, Nevins, & Levine 2012), and spoken production (e.g. Wilson 2006, Peperkamp, Skoruppa, & Dupoux 2006, Baer-Henney 2015).

The ways in which participants are exposed to or trained on the artificial language also vary along many dimensions. The words of the language may be associated with referents, usually pictures (e.g. Peperkamp, Skoruppa, & Dupoux 2006, White 2013), or simply presented auditorily without referents (e.g. Finley 2008). An alternation may be taught by presenting full paradigms, such as when singular forms are followed immediately by their corresponding plural forms (e.g. Pycha et al. 2003, White 2013), or by presenting stems and affixed forms in random order, unorganized by paradigm (e.g. Baer-Henney & van de Vijver 2012). The proportion of training items that are fillers (items not participating in the crucial phonological pattern) can vary, from two thirds in Greenwood 2016 to half in Wilson 2006 or White 2013 to about 5% in Shapp 2012. In some experiments' training phases, participants are obliged to produce words of the artificial language out loud (e.g. Moreton 2008, 2012, Sturman 2018) or are prompted more or less strongly to do so (e.g. Wilson 2006, White 2013, Myers & Padgett 2014) while in other cases there are no instructions or encouragement to do so (e.g. Finley 2008). Training phases may include a section with feedback to the participant (e.g. Pycha et al. 2003) or not (e.g. Finley 2008). Instructions to participants also differ: some just tell participants they will be listening to words of an unfamiliar language (e.g. Finley 2008, Baer-Henney & van de Vijver 2012) while some tell participants they should figure out a process, such as how to form plurals in the language (e.g. Pycha et al. 2003). Some of these factors are associated with more or less explicit learning (Moreton & Pertsova 2016); AGL experiments generally seek to foster more implicit learning of the phonological patterns being tested so that the type of learning participants are engaging in more closely resembles native language acquisition. I further discuss how much AGL studies can reveal about natural language acquisition in the next section. For a review of AGL work testing for synchronic learning biases, see Moreton & Pater 2012a,b.

While experiments investigating bias are generally straightforwardly categorized as natural language studies or AGL studies, at least one experiment, Skoruppa & Peperkamp 2011, is arguably a mix of the two. This study taught native French speakers artificial dialects of French featuring real French words modified to exhibit vowel (dis)harmony. The experiment tested the learning of artificial phonological patterns absent from the participants' native language, as standard French does not have vowel harmony, so in this sense it is an AGL study. The patterns were embedded in the participants' native language, though, which may have made the experiment more naturalistic.

1.1.2 What Kind of Learning Do Phonological Experiments Tap Into?

Because they test the learning of different things, natural language studies and AGL experiments likely shed light on different learning mechanisms. Natural language studies tap into the accumulated knowledge participants have acquired about their language over the entire course of their lives preceding the experiment. They test what participants have learned from real language input delivered naturally. Thus discrepancies between the generalizations present in the natural language input and the generalizations participants demonstrate they have extracted from that input should represent the effect of biases that influenced participants' L1 learning, that is, biases that are active in natural phonological acquisition. AGL experiments, on the other hand, test the learning of novel patterns taught through brief, non-naturalistic exposure in a laboratory setting. Thus one must be cautious in interpreting AGL results: just because a bias emerges in an AGL study does not mean it is active in language acquisition, and, similarly, just because a bias does *not* emerge in an AGL study does not mean it is not active in acquisition. Most AGL studies are conducted on participants who are adults and thus past the age of L1 acquisition, so biases that are specific to L1 acquisition may not come out in their performance. (Even when participants are infants, exposure to the artificial language is still brief and non-naturalistic.)

Additionally, if participants in AGL experiments are using learning mechanisms other than those used in L1 acquisition, these experiments may not tell us much about what factors shape natural phonological learning. Moreton & Pertsova (2016:289) propose that artificial grammar learning's closest natural language counterpart is in fact second-language (L2) acquisition, a process that often involves both implicit and explicit learning mechanisms. Thus insofar as L2 phonological acquisition is also of interest to linguists, AGL studies are worth conducting. However, the goal of most AGL studies on synchronic learning biases seems to be uncovering the biases of the language faculty when faced with the task of learning a native language. Presumably it is through L1 acquisition that synchronic biases could shape the phonological typology, and explaining typological asymmetries is often the motivation for experimental investigations of learning bias (e.g. Finley 2008, Skoruppa & Peperkamp 2011, Shapp 2012, White 2013, Kimper 2016).

Another question that arises in the study of synchronic learning biases is whether the biases that emerge in experiments are specific to language or domain-general. Substantive bias, by definition, must be linguistic in nature because it refers to the articulation and perception of speech sounds (Moreton 2012). The other major type of synchronic bias investigated in the literature, complexity bias, need not necessarily be linguistic, though. It seems intuitively right that more complex patterns, whether in language or some other domain, would be more difficult to learn, and indeed patterns that require more dimensions to define have been shown to be harder in both visual concept learning and phonological learning (Shepard, Hovland, & Jenkins 1961, Moreton, Pater, & Pertsova 2017). Thus at least some types of complexity bias are surely domain-general, although potentially other types could be more specific to language (Moreton 2012).

Domain-general cognitive biases should apply to linguistic learning as well as to any other type of learning. For instance, Wilson (2003) found a learning advantage for phonological patterns in which the choice of a suffix allomorph was driven by nasal assimilation or dissimilation over patterns in which the suffix allomorph was determined more arbitrarily. He favored attributing this result to domain-general biases toward identity and similarity. Carpenter (2010) proposes that both domain-general and language-specific learning mechanisms are active in language acquisition but language-specific mechanisms only apply to the learning of phonetically natural patterns, making these patterns easier to acquire. If, due to their many non-naturalistic properties, AGL experiments did not lead participants to engage in linguistic learning, they would not be able to capture language-specific biases. Instead, they would only detect domain-general biases that are still valid for whatever type of learning participants are actually engaging in. This does not seem implausible given the scarcity of AGL results robustly supporting substantive (and thus unequivocally language-specific) bias (see Section 1.2.2), but it may be too strong a position. As outlined earlier, AGL experiment designs vary in many ways, and some designs may be able to elicit linguistic learning and uncover language-specific biases.

Giving the problems with drawing conclusions about L1 phonological acquisition from AGL experiments, the best way to investigate synchronic learning biases is to use both natural language and AGL studies. When both types of studies yield the same effect, there is stronger evidence that the effect in question reflects a genuine bias (domain-general or not) that is active in linguistic learning. Additionally, if AGL experiments can be shown to corroborate natural language studies fairly consistently, it strengthens the argument that AGL is a legitimate way to test for learning biases. This in turn renders more credible claims of synchronic learning biases based only on evidence from AGL and not on evidence from natural language studies. Some AGL studies do find effects of substantive bias, so it seems that this methodology must be able to tap into linguistic learning at least sometimes. In the following sections, however, I will show that there is less evidence for substantive bias from AGL than has been claimed, so caution remains warranted. AGL results should still be considered together with results from natural language studies to obtain the most accurate picture of the biases that constrain phonological learning.

1.2 Experimental Evidence for Synchronic Learning Biases in Phonology

In the following sections, I review the experimental literature on synchronic learning biases. I begin with studies that have found effects of complexity bias. Then I turn to substantive bias and discuss its rival, channel bias, and the definition of naturalness. My review of the substantive bias literature considers studies of three types: those that purport to have found effects of substantive bias but which I argue do not control for complexity and therefore may only have found effects of complexity bias, those that appear to have found genuine substantive bias effects, and those that failed to find effects of substantive bias. Unless otherwise specified, all studies were conducted on native speakers of English.

1.2.1 Complexity Bias

A number of studies have uncovered evidence for complexity bias by finding that learners acquire featurally simpler patterns better than patterns that are featurally more complex. In an AGL experiment, Pycha et al. 2003 tested the learning of three patterns: vowel harmony, vowel disharmony, and an arbitrary vowel alternation. In the harmony pattern, the suffix vowel agreed in backness with the stem vowel, surfacing as [ɛ] after the front vowels [i ɪ æ] and as [ʌ] after the back vowels [u ʊ a]. In the disharmony pattern, the suffix vowel disagreed in backness with the stem vowel, surfacing as [ɛ] after back vowels and as [ʌ] after front vowels. In both cases, the value of the [back] feature of the suffix vowel was predictable from the value of the [back] feature of the stem vowel. In the arbitrary pattern, the suffix vowel was [ɛ] after [i æ ʊ] and [ʌ] after [u a ɪ]. In this pattern, the value of the [back] feature of the suffix vowel could not be predicted from the value of the [back] feature of the stem vowel. Any formalization of this pattern will be more complex than the formalization of the harmony and disharmony patterns. Pycha et al. found that both the backness harmony and disharmony patterns were learned better than the arbitrary pattern, supporting complexity bias. Similarly, Skoruppa & Peperkamp (2011) found that French speakers learned artificial phonotactic patterns of vowel rounding harmony and disharmony better than a mixed harmony/disharmony pattern that depended on more than just the [round] feature.

In two AGL experiments, Moreton (2008) tested for synchronic learning bias by comparing the learnability of different phonotactic patterns. In Experiment 1, the two patterns were a height-height dependency in which vowels within a word agreed in height (e.g. [t**i**d**u**]) and a height-voicing dependency in which vowels were high preceding voiced stops and low preceding voiceless stops (e.g. [t**id**i]). The two patterns are equally phonetically natural in that the phonetic precursors to the pattern are equally robust in both cases. However, participants learned the height-height pattern better than the height-voicing pattern, as evidenced by their ability to choose the pattern-conforming word in a two-alternative forced choice task. In Experiment 2, the two patterns were the height-voicing dependency and a voicing-voicing dependency in which consonants in a word agreed in voicing (e.g. [**g**i**d**i]). Unlike the height-voicing dependency, the voicing-voicing dependency does not have phonetic precursors, yet it was learned better than the height-voicing dependency, though not as much better as the natural height-height dependency was. Moreton concludes that the effects in both experiments are due to a type of complexity bias whereby single-feature dependencies (in this case, those that relate two instances of the feature [high] or the feature [voice]) are favored over multiple-feature dependencies (in this case, the pattern that relates one instance of [high] and one instance of [voice]).

In a series of follow-up experiments, Moreton (2012) obtained further evidence supporting a learning advantage for intradimensional (single-feature) phonotactic patterns over interdimensional (multiple-feature) ones. His results showed that it was truly relating instances of the same feature that mattered and not simply relating segments on the same tier (vowel-vowel or consonant-consonant dependencies vs. vowel-consonant dependencies). Moreton's studies demonstrate that featural complexity is not just a matter of the number of instances of features needed to express a pattern (Chomsky & Halle 1968) but also encompasses whether the instances are of the same feature or of different features.

Baer-Henney & van de Vijver (2012) explored the effects of a number of factors on the learning of artificial morphophonemic alternations. Two of the patterns they taught their German-speaking participants were phonetically unnatural but differed in complexity (or, as the authors put it, locality). In both artificial languages, the plural suffix had a front allomorph [-y] and a back allomorph [-u]. In the local pattern, the choice of allomorph was conditioned by the tenseness of the stem vowel while in the non-local pattern it was conditioned by the sonorancy of the stem-initial consonant (all stems had the shape CVC). Thus in the local pattern participants had to learn a constraint of the form *[α tense][β back] (the stem vowel and the suffix vowel being adjacent on the vowel tier) while in the non-local pattern they had to learn a constraint like *[+son]VC[+back]. After an exposure phase, participants had to produce the plural forms for novel singular stems. Their performance showed that the simpler local pattern was learned better than the more complex non-local pattern.

Complexity bias effects have also been found in AGL experiments carried out on infants. Cristià & Seidl (2008) conducted an AGL study examining whether infants use features in phonotactic learning. In Experiment 1, 7-month-old infants were exposed to words beginning either with stops or nasals (the natural condition) or with nasals or fricatives (the arbitrary condition). The legal onsets in the natural condition formed a natural class defined by [–cont] while the legal onsets in the arbitrary condition did not form a natural class definable by conventional phonological features. The infants were then tested on words beginning with novel stops (illegal in the arbitrary condition) and words beginning with novel fricatives (illegal in the natural condition). If in the test phase infants differentiated stop- and fricative-initial words, this would show they had learned the phonotactic restriction of the language they had been exposed to. Infants in the natural condition listened longer to fricative-initial test words (a novelty preference), thereby demonstrating learning, but infants in the arbitrary condition did not distinguish between legal and illegal test words, thus showing no evidence of learning. In Experiment 2, 7-month-old infants were simply exposed to stop-initial or fricative-initial words and tested on both. Infants in both groups distinguished the type of onset they had heard in training from the type of onset they had not heard, listening longer to legal onsets (a familiarity preference).

Cristià & Seidl concluded that infants can make use of features in phonotactic learning, whether these features are innate, abstract phonological features like [–cont] or concrete phonetic features infants induce. I consider their results to support complexity bias in either case.[1] In the natural condition of Experiment 1, infants could have learned the phonotactic constraint *[+cont] to differentiate illegal onsets from legal ones while in the arbitrary condition they would have had to learn a constraint like *[–cont, –son] to categorize stops as illegal onsets without doing the same to nasals, which were licit onsets. Thus the natural condition was simpler than the arbitrary condition, and only in the natural condition did infants demonstrate learning. In Experiment 2, the exposure phase lacked nasal onsets, so infants could simply learn *[+cont] or *[–cont] to differentiate legal and illegal onsets. The conditions were equally simple, and infants showed learning in both.

Even if infants were using something more physical than abstract features like [cont], I believe the study still reveals an effect of complexity bias. In the natural condition of Experiment 1, infants were sensitive to the illegal fricative onsets, which are more acoustically salient than the legal onsets, which were stops and nasals. In the arbitrary condition, infants were not able to group the more acoustically salient fricatives and the less acoustically salient nasals to the exclusion of the also less acoustically salient stops, and consequently in the test phase they failed

---

[1] This dissertation will not be concerned with whether phonological features are innate or induced. I do not assume that features are innate, as in Chomsky & Halle’s (1968) theory and as argued in Hale, Kissock, & Reiss 2006, but neither do I claim that features must be learned or emergent (for this argument, see Mielke 2004). While the question of what is innate and what is learned is relevant to the study of phonological learning biases (see Section 1.2.2.1), for the purposes of this dissertation I set aside the question of whether features are innate.

to distinguish legal and illegal onsets. Another artificial phonotactic learning study with infants that yields evidence for complexity bias is Saffran & Thiessen 2003, which I discuss further in Section 1.3.1. These studies show that phonological learning is constrained by complexity bias in infants as well as adults.[2]

Thus far, I have cited complexity bias effects found in AGL studies, but evidence for complexity bias also comes from natural language studies. Hayes et al. (2009) investigated bias in the learning of constraints governing the alternation of the dative suffix in Hungarian. Hungarian exhibits backness harmony in suffixes, and the dative suffix alternates between back [-nɔk] and front [-nɛk]. Hungarian stems with certain patterns of front and back vowels take only one form or the other, but other kinds of stems appear with both, to different degrees of frequency (e.g. stems ending in coronal sonorants tend to, but do not always, take the front form [-nɛk]). Hayes et al. identified patterns in stems that correlated with a certain allomorph of the dative suffix. They then conducted a wug test in which Hungarian speakers judged the goodness of each allomorph when suffixed to nonce stems in order to determine the extent to which speakers had learned the patterns in the lexicon. They found that, compared to simple natural constraints, more complex natural constraints governing dative allomorph selection were underlearned relative to their strength in the lexicon. This result furnishes support for complexity bias in a natural language study.

As demonstrated by the studies reviewed here, complexity bias has been found in natural language learning and artificial language learning and in both adults and infants. Complexity can also manifest in different ways. In terms of featural complexity, a pattern is more complex if it

---

[2] While promising, these infant study results should be taken with a grain of salt. Cristia (2018) conducted a meta-analysis of sixteen infant phonotactic learning studies, including Cristià & Seidl 2008, and found an average effect size of zero, casting doubt on whether infants actually demonstrate phonotactic learning in AGL experiments.

requires more features to express (e.g. *[−cont, −son] vs. *[+cont] in Cristià & Seidl 2008) or if it is interdimensional rather than intradimensional (e.g. *[+high][−voice] vs. *[+high][−high] (on the vowel tier) in Moreton 2008, 2012). Non-locality also appears to contribute to pattern complexity (e.g. Baer-Henney & van de Vijver 2012). For further discussion of the types of formal complexity and a broader review of complexity bias findings, see Moreton & Pater 2012a.

1.2.2 Substantive Bias

1.2.2.1 An Alternative to Substantive Bias: Channel Bias

Substantive bias, a learning bias that favors phonetically natural phonological patterns, is only one hypothesis for why the phonological typology displays asymmetries rooted in phonetic naturalness. Its major competitor in the debate over the cause of these asymmetries is the diachronic explanation, sometimes called channel bias (Moreton 2008). The channel is the channel of language transmission, which takes place through speech production and perception (Moreton & Pater 2012a). Channel bias is not a cognitive bias of the learner's language acquisition system but instead refers to the misproduction and misperception that disproportionately affect certain sounds, contrasts, or structures.

The nature of channel bias is determined by articulatory difficulty and perceptibility. For example, since voiced obstruents are more difficult to produce in word-edge position (Westbury & Keating 1986), speakers may occasionally devoice, or consistently partially devoice, word-edge voiced obstruents, causing listeners to perceive the obstruents as voiceless and mislearn them as underlyingly voiceless. The effect of channel bias is to drive the imperfect transmission of languages over time, and phonetic factors ensure that the resulting changes to phonological grammars tend to give rise to increasingly natural phonological systems (Ohala 1992, Blevins 2004). It is thus the effect of articulatory and perceptual ease on language change that led to the present typology.[3]

A common objection to embracing both channel bias and substantive bias is that it is unparsimonious (Ohala 1974, Hale & Reiss 2000, Blevins 2004, Greenwood 2016): if channel bias can filter out all the phonological patterns that do not (or only rarely) arise, it suffices, and we should not additionally postulate a cognitive bias against acquiring these phonological patterns. Not all proponents of channel bias consider it the sole force shaping phonological typology, to the exclusion of all synchronic bias. Moreton (2008:88, 118) argues that channel bias cannot explain all aspects of the typology (cf. Hayes & Steriade 2004:26-27) and therefore sees a role for both types of bias, including synchronic bias rooted in phonetic naturalness. Indeed, he claims that it would be unparsimonious to assume there are no effects of substantive bias (though later Moreton & Pater 2012b cast doubt on the existence of substantive bias given the dearth of experimental evidence for it). It is possible that both substantive bias and channel bias contribute to the general naturalness of phonological systems cross-linguistically, though it is of course very difficult to determine which is responsible for a given typological fact (Moreton 2008).

---

[3] Ohala (1992) maintains that the locus of sound change is perception; all sound change originates in hypo- or hypercorrection on the part of the listener. His position still leaves room for articulatory difficulty to shape sound change, however, because listener misperception can result from a failure to compensate for speaker variation. This variation in production surely arises at least in part due to articulatory difficulty. Ohala also disputes the notion that sound change is the result of speakers' intention to improve or simplify the language (the teleological view); Blevins (2004) shares this non-goal-oriented conception of language change. I agree that natural phonological patterns are emergent, not sought by speakers. Ohala perhaps goes further by stating that he does not believe sound change makes languages' sound patterns better or simpler (e.g. easier to produce, perceive, learn, or process). It seems to me that at the very least sound change *can* do this: if articulatory difficulty causes speaker variation which is misconstrued by listeners, the ultimate result could be the elimination of an articulatorily difficult structure (e.g. final voiced stops), meaning that sound change has eased production.

Also at issue is the question of how substantive bias, assuming it exists, affects phonology: is it built into Universal Grammar or is it a constraint on learning that derives from knowledge that is learned? The argument from parsimony against substantive bias often seems to suppose that the bias is a property of innate knowledge (Ohala 1974, Hale & Reiss 2000); for example, Ohala (1974) rejects that newborns simply know that it is better for voiced obstruents to be devoiced. Since the advent of Optimality Theory (Prince & Smolensky 1993/2004), this innate knowledge has often thought to take the form of a universal constraint set, provided by Universal Grammar, which contains some constraints but not others, thus encoding substantive bias. This is the conception of substantive bias as innate that is criticized in, for instance, Greenwood 2016 and Hale & Reiss 2000.

The existence of substantive bias is not inextricably tied to its being innate, though. The phonetic knowledge that leads learners to prefer certain phonological patterns and systems over others could itself be learned (Hayes & Wilson 2008:425, White 2013:167). Infants are accruing phonetic experience, both perceptual and articulatory, even as they are in the process of acquiring the phonological system of their native language. It seems conceivable that the phonetic knowledge they accumulate could interact with the acquisition of their native phonological system (see Hayes 1999 for a proposal), as well as later learning, as in the case of L2 acquisition (Moreton & Pertsova 2016).

The question of whether the phonetic knowledge that biases phonological learning is learned or innate is beyond the scope of this dissertation. The experiments I present test for effects of such a bias, not for its origin. For now, it is enough to point out that rejecting that knowledge of phonetic naturalness is innate (perhaps built into a universal constraint set) does not necessarily mean rejecting a synchronic, cognitive learning bias rooted in phonetic naturalness.

1.2.2.2 Defining Phonetic Naturalness

In my review of experiments testing for substantive bias, I will show that many experiments fail to control for complexity when manipulating naturalness to try to detect a naturalness bias effect. Part of the reason so many substantive bias studies contain confounds is that the definition of naturalness varies through the literature and is also sometimes left vague. Wilson (2003) describes his study as an investigation of naturalness even as he acknowledges using the term loosely. Later, in discussing the nature of the cognitive bias that made a nasal assimilation rule easier to learn than an arbitrary rule, he refers to an identity relation between the trigger and the target's [+nasal] features, which recalls Moreton (2008, 2012)'s intradimensional advantage, and to the similarity between the trigger and the target, which are both also [+son] and [–cont] in addition to [+nasal]. This identity and similarity are lacking in the arbitrary rule, but both, especially identity, seem to fall under formal complexity more than phonetic naturalness.

Typology also seems to play an implicit role in what makes Wilson's nasal assimilation and dissimilation patterns more natural than their arbitrary counterparts: he cites languages that exhibit the assimilation and dissimilation patterns but notes that the arbitrary pattern is unattested. Carpenter (2006) reads naturalness directly off the typology in explicit fashion; the patterns she considers natural are those which are attested, and unattested patterns are deemed unnatural. In a more mixed approach, Hayes & White (2013) allow typological frequency to be a sufficient condition for naturalness, defining natural patterns as those which are typologically common *or* which promote ease of articulation or perception. To qualify as unnatural, patterns had to be largely unsupported by the typology or have little phonetic benefit. To be sure, typological frequency and phonetic naturalness generally go hand-in-hand; the substantive bias enterprise assumes that phonological systems tend to be phonetically natural and seeks to determine whether this is due to a cognitive learning bias. Alternatively, it could be due to the way languages change over time (Blevins 2004). In either case, as Greenwood (2016) points out, typological frequency is the *result* of naturalness; it is not what makes a pattern natural. Thus typological frequency is not evidence that a pattern is natural, and I do not consider it a sufficient condition for naturalness.

Pycha et al. 2003 is a relatively early study that uses a narrower definition of naturalness than either Wilson 2003 or Hayes & White 2013. They consider a pattern natural only if it has phonetic precursors in the speech signal (cf. Moreton 2008). For example, vowel harmony, one of the patterns they test, is natural because it can develop from vowel-to-vowel coarticulation. Pycha et al.'s naturalness criterion seems to be particularly narrow, apparently excluding promoting articulatory ease or providing perceptual benefits unrelated to a phonetic precursor, as in the case of dissimilation. Hayes & White's (2013) phonetic criterion for naturalness is broader, encompassing ease of articulation and perception. Similarly, Finley & Badecker (2012:168) understand phonetic naturalness to mean "maximiz[ing] perceptual salience and articulatory ease."

Varying definitions of naturalness can sometimes lead to conflicting classifications for the same type of phonological pattern. For instance, both Pycha et al. (2003) and Skoruppa & Peperkamp (2011) consider vowel disharmony to be unnatural and compare its learnability with its natural counterpart, vowel harmony. For Pycha et al., vowel disharmony is unnatural because it does not arise from phonetic precursors, and for Skoruppa & Peperkamp it is unnatural because it is typologically rarer and has no clear phonetic motivation. Typological rarity aside, if vowel disharmony is not phonetically motivated it is appropriate to treat it as unnatural.[4] Baer-Henney (2015) notes, however, that dissimilation is sometimes deemed unnatural, as with vowel disharmony in Pycha et al. 2003, and sometimes deemed natural, as with nasal dissimilation in Wilson 2003 (though Wilson's definition of naturalness is very broad and does not refer to phonetic motivation). Is it possible for dissimilatory patterns to be phonetically grounded too? Pycha et al. (2003) recognize that the argument that vowel harmony facilitates perception by making vowels more predictable from other vowels would work for vowel disharmony too, meaning that disharmony would also facilitate perception. From this perspective, disharmony should be as natural as harmony, though Pycha et al. do not opt for this interpretation of the equal learnability of harmony and disharmony in their study. Others have also suggested that dissimilatory patterns are natural insofar as they provide perceptual benefits (Johnson 1973, Frisch 2004, Bye 2011; see also Ohala 1992 for discussion of dissimilation as a natural sound change). Thus despite its relative rarity compared to assimilation, dissimilation may not be phonetically unmotivated or unnatural, reinforcing that typology frequency alone should not be a measure of naturalness.

My own view is that a phonological pattern qualifies as natural if it can be shown to be phonetically motivated. The phonetic motivation may be articulatory or perceptual, in the vein of Finley & Badecker 2012 and Hayes & White 2013. Typological frequency is interesting to

---

[4] At least, it is appropriate to consider vowel disharmony non-natural, that is, unmotivated by phonetic factors. Beguš (2018) distinguishes natural, unmotivated, and unnatural patterns, where unnatural patterns are not just phonetically unmotivated but actually operate against phonetic naturalness, by *decreasing* ease of articulation and perception. While recognizing the validity of this distinction, I do not emphasize it here and use the term unnatural to cover any pattern that is not phonetically motivated.

consider but does not itself make a pattern natural or unnatural. My definition is in the vein of Ohala 1974, which equates naturalness with being rooted in universal articulatory or auditory factors. Ohala recognizes that phonologists sometimes use “natural” in a broader sense to designate phonological patterns that the cognitive system handles better, but his definition of naturalness does not extend as far, and neither does mine. Thus in my review below I do not include naturalness studies that test the learnability of patterns that are deemed “natural” for more formal, rather than concretely phonetic, reasons. Such patterns include directional (vs. majority rules) vowel harmony (Finley & Badecker 2008) and stem-final alternations in polysyllables (vs. monosyllables) (Becker, Nevins, & Levine 2012, Becker, Clemens, & Nevins 2017).

1.2.2.3 Substantive Bias Effects Confounded with Complexity Bias

In the synchronic learning bias literature, there are numerous studies that claim to have uncovered evidence for substantive bias but which do not adequately control for complexity, such that their results may in fact reflect complexity bias effects. Given the relative scarcity of indisputable substantive bias results (Moreton & Pater 2012b), the conservative approach would be to interpret these studies as revealing complexity bias, thus providing even more support for this kind of bias. In any case, when naturalness and complexity are confounded, a difference in learnability cannot be taken as reliable evidence for substantive bias. In this section, I review studies that fall into this category.

An early AGL study purporting to test for naturalness bias is Wilson 2003. As mentioned above, Wilson’s definition of naturalness is loose, and in the end he concludes that the bias he finds has to do with identity and similarity, but the study is sometimes cited as supporting substantive bias (e.g. Finley 2008, Hayes et al. 2009). Wilson's artificial languages consisted of CVCV stems and a suffix with two allomorphs, [-na] and [-la].[5] He exposed participants to suffixed forms exhibiting a certain suffix alternation pattern and then tested them on stems bearing correct and incorrect suffix allomorphs. If participants accepted new pattern-conforming suffixed forms more than new nonconforming forms, this demonstrated they had learned the alternation.

In Experiment 1, Wilson tested the learning of nasal assimilation (the suffix surfaces as [-na] after stems containing a nasal and as [-la] otherwise) and a random alternation (the suffix surface as [-na] after stems containing a dorsal consonant and as [-la] otherwise). In Experiment 2, he compared nasal dissimilation (the suffix surfaces as [-na] except after stems containing a nasal, when it surface as [-la]) to the same random alternation. Participants were able to learn both nasal assimilation and nasal dissimilation but not the random pattern. The assimilation and dissimilation patterns may be more natural than the arbitrary one, but they are also clearly simpler. Though Wilson claims that the two patterns in each experiment are formally matched, this is only true by a feature-counting metric. The assimilation and dissimilation patterns are both intradimensional (Moreton 2012), relating two instances of the same feature, [nasal], while the random pattern is interdimensional, relating one instance of [dorsal] and one instance of [nasal]. Thus, as noted in Gordon 2016, naturalness and complexity are confounded, and Wilson's results may simply be due to complexity bias (as his own interpretation suggests). Wilson also claims that nasal assimilation and dissimilation are learned equally well, though he does not directly compare the two with a statistical test. This might constitute a null substantive bias result if one

---

[5] Participants were exposed only to suffixed forms, not paradigms of bare stems and suffixed forms, so they may or may not have arrived at the morphological parse that underlay the design of the artificial languages. If they did not, they presumably treated the task as phonotactic learning rather than alternation learning.

of the alternations could be shown to be more phonetically motivated than the other, but this has not been done (indeed, in Wilson's framing, the two patterns are on equal footing).

Peperkamp, Skoruppa, & Dupoux (2006) taught French speakers "natural" allophonic distributions of voiced and voiceless obstruents (either stops or fricatives had predictable voicing, depending on the subcondition) and "unnatural" allophonic distributions relating arbitrary consonants. The distributions were demonstrated with intervocalic alternations: in the natural languages, the relevant obstruents voiced intervocalically while a sample alternation in the unnatural languages was /g/ surfacing as [f] intervocalically. The test phase was a production task in which participants had to apply the alternations they had been trained on. Performance was better for the natural languages than for the unnatural languages, and the authors claim to have found an effect of naturalness bias. However, the unnatural languages are more complex than the natural ones, since the natural voicing alternations can be captured with a natural class generalization (e.g. *V[−voice]V) while the arbitrary alternations cannot. Thus what Peperkamp, Skoruppa, & Dupoux uncovered may well be a complexity bias effect.

Skoruppa, Lambrechts, & Peperkamp (2011) compared the learning of alternations relating sounds differing in one (e.g. [p]~[t]), two (e.g. [p]~[s]), or three (e.g. [p]~[z]) features. They consider the phonetic distance between two alternating sounds another factor that can make a pattern more or less natural, with naturalness decreasing as phonetic distance increases. They tested French-speaking participants with a production task similar to that of Peperkamp, Skoruppa, & Dupoux 2006 and found that the one-feature alternations were learned better than the two- and three-feature alternations, which did not differ from each other. While Skoruppa, Lambrechts, & Peperkamp interpret this result as supporting naturalness bias and White (2013) cites it that way, the one-feature alternations are also simpler than the other alternations by virtue of the number of features required to express the phonological change. Once again, naturalness and complexity are confounded, so the result could just as easily represent a complexity bias effect. Indeed, Moreton & Pater (2012a) cite Wilson 2003, Peperkamp, Skoruppa, & Dupoux 2006, and Skoruppa, Lambrechts, & Peperkamp 2011 as studies yielding evidence for complexity bias.

Another artificial alternation learning study that confounds naturalness and complexity is Baer-Henney & van de Vijver 2012. This experiment found a clear complexity bias effect based on locality discussed in Section 1.2.1, but Baer-Henney & van de Vijver also compared the learnability of two local patterns, one of which they considered natural. The unnatural pattern was the one in which the suffix allomorph, front [-y] or back [-u], was conditioned by the tenseness of the stem vowel. The natural pattern was simply backness harmony: the suffix agreed in backness with the stem vowel (e.g. [l**ɪ**t**-y**]). German speakers learned the natural backness harmony pattern better than the unnatural tenseness-backness pattern, as evidenced by a production task, and Baer-Henney & van de Vijver take this as evidence for substantive bias. As in Wilson 2003, though, the natural pattern in this study is intradimensional while the unnatural pattern is interdimensional. The natural pattern is therefore also simpler, so the fact that participants mastered it better may constitute an effect of complexity bias.

In a related study, Baer-Henney, Kügler, & van de Vijver (2015) complicate the picture further. They taught German speakers the same backness harmony and tenseness-backness pattern as in Baer-Henney & van de Vijver 2012, but they varied amount of exposure to the patterns (two vs. three training blocks) and amount of exceptionality (15% vs. 35%). That is, in conditions with 15% exceptionality, 15% of suffixed forms were “disharmonic” with respect to the pattern being taught. Additionally, the authors investigate the influence of L1 phonotactics on learning of the artificial alternations. They claim that tenseness, but not backness, is a relevant feature in German phonotactics (e.g. lax vowels can only occur in closed syllables), so the tenseness-backness pattern is supported by the participants' native phonotactics in a way that backness harmony is not. (This was also true in Baer-Henney & van de Vijver 2012 but is not considered as a factor in that study.) Manipulating naturalness and L1 phonotactic relevance simultaneously means that neither is controlled for: backness harmony is natural but unsupported by German phonotactics while the tenseness-backness pattern is unnatural but supported by German phonotactics.

Based on the results of the production task, the general picture was that participants learned backness harmony better than the tenseness-backness pattern only with shorter exposure and greater exceptionality. Otherwise, they learned the tenseness-backness pattern better than backness harmony, usually significantly. This suggests an effect of substantive bias only when the input is poor, but as in Baer-Henney & van de Vijver 2012, this could be an effect of complexity bias. The unnatural (and more complex) tenseness-backness pattern may have been learned better than backness harmony in the other exposure and exceptionality conditions because it is supported by German phonotactics, but in that case these results are hard to reconcile with those of Baer-Henney & van de Vijver 2012, in which backness harmony was still learned better than the tenseness-backness pattern with exceptionless input.

I am aware of one natural language study that claims to have found evidence for substantive bias but may only have found evidence for complexity bias. Becker, Ketrez, & Nevins (2011) used a wug test to probe Turkish speakers' knowledge of phonological properties that make stems more or less likely to undergo laryngeal alternations. In general, Turkish has final devoicing, so the final voiceless consonant of a bare stem may or may not voice when suffixation puts it in an intervocalic context. Moreover, certain stem types are more likely to "voice intervocalically." Becker, Ketrez, & Nevins uncovered four patterns in the Turkish lexicon that were predictive of rate of alternation (i.e. intervocalic voicing) of stem-final voiceless stops: 1) the final stops of polysyllabic stems alternate more than those of monosyllabic stems; 2) of the places of articulation, stem-final labials and velars alternate the most while stem-final coronals alternate the least; 3) stem-final stops preceded by a high vowel alternate more than those preceded by a non-high vowel; and 4) stem-final stops preceded by a back vowel alternate more than those preceded by a front vowel. They consider the word length and place of articulation regularities natural and the vowel height and backness regularities unnatural, but these classifications are based on typological frequency rather than phonetic motivation.

To test whether Turkish speakers have internalized these patterns, Becker, Ketrez, & Nevins orthographically presented participants with nonce bare stems (e.g. <fet>, representing [fet^h]) and asked them to choose between two auditorily presented suffixed forms, one in which the stem-final consonant alternated and one in which it did not (e.g. [fed-i] and [fet^h-i]). The Turkish speakers reproduced the word length and place of articulation patterns but not the vowel height and backness patterns, which the authors construe as support for naturalness bias. Since they do not justify the naturalness of the word length and place of articulation effects on phonetic grounds, I do not consider this a substantive bias result. Additionally, the natural patterns are simpler than the unnatural patterns; Becker, Ketrez, & Nevins consider the word length and place of articulation patterns first-order phonotactics while the vowel height and backness patterns are considered second-order phonotactics, since they relate a feature of a vowel to the behavior of a following consonant. Consequently, the fact that Turkish speakers have learned the word length and place of articulation patterns from the lexicon but not the vowel height and backness patterns may be an effect of complexity bias.

The studies I have reviewed above all deal with phonological alternations. There are also phonotactic learning studies in which naturalness is confounded with complexity, but I will discuss these in Section 1.3, which focuses specifically on the literature on substantive bias in phonotactic learning. What I have shown in this section is that some past studies that appear to test for naturalness bias do not compare the learning of clearly phonetically motivated patterns with phonetically unmotivated ones because the authors' definition of naturalness was not closely tied to phonetic grounding (Wilson 2003, Becker, Ketrez, & Nevins 2011). All the studies I have discussed here fail to control for complexity when choosing natural and unnatural patterns to compare. Sometimes this is because intra- vs. interdimensionality (Moreton 2012), as opposed to simple feature counting, is not taken into account as a possible source of complexity (Wilson 2003, Baer-Henney & van de Vijver 2012). Because complexity bias cannot be ruled out as the explanation for the effects uncovered in these studies, they cannot be considered reliable evidence for substantive bias, and the number of experiments that support substantive bias is reduced.

I have been reinterpreting substantive bias effects as complexity bias effects by pointing out how the natural patterns are more simply expressed in features, whether due to intradimensionality or the number of features required, than their unnatural counterparts. I argued that in many studies naturalness is confounded with complexity. It is worth considering, however, how complexity might itself be confounded with naturalness. The feature system we use to express phonological generalizations is designed to capture natural patterns more simply than unnatural ones (Moreton & Pater 2012b:709). Thus it is conceivable that a natural pattern might be featurally simpler than an unnatural pattern because of a bias built into the feature system and not because the natural pattern is actually simpler from the point of view of human cognition. Pushing this line of reasoning to its extreme, one could imagine reinterpreting complexity bias effects as substantive bias effects: the simple pattern can be expressed more economically in features precisely because it is more natural and the feature system was designed to express natural patterns with fewer features. The more we can be sure the features used in phonological analysis are actually used by the phonological learning mechanism and in speakers' grammars, the less of a concern this property of the feature system is. On this point, studies that find evidence for the psychological reality of classical distinctive features (e.g. Fromkin 1971, Stemberger 1983, Durvasula & Nelson 2018; see Mowrey & MacKay 1990 and Frisch & Wright 2002 for discussion) are encouraging. Importantly, it is also still possible to compare the learning of natural and unnatural patterns that are truly of equal formal complexity according to the standard system of features, whatever its biases. That is, the feature system, despite its design, can still express certain unnatural, phonetically unmotivated patterns simply. In the next section, I turn to experiments investigating substantive bias that compare natural and unnatural patterns matched for complexity under the usual feature system. I continue to focus on alternation learning studies, saving discussion of phonotactic learning for Section 1.3.

1.2.2.4 Robust Substantive Bias Effects

Many studies have tested for naturalness bias in phonological learning while avoiding the confound of complexity. Of these, some have found convincing evidence for substantive bias. Hayes et al.'s (2009) natural language study with Hungarian speakers, discussed in Section 1.2.1 on complexity bias findings, also found underlearning of unnatural constraints governing the choice of dative suffix allomorph relative to their strength in the lexicon. The unnatural constraints that were underlearned were not more complex than certain natural constraints that were learned to the degree one would predict from the lexicon. Thus Hayes et al. 2009 found evidence for substantive bias as well as complexity bias.

Shapp (2012) compared the learnability of sibilant (anteriority) harmony and voicing harmony, which she taught in an AGL experiment using alternating plural suffixes whose initial consonants agreed in anteriority (e.g. [ta**z**i-**s**a], [ta**ʒ**i-**ʃ**a]) or in voicing (e.g. [ma**t**i-**p**a], [ma**d**i-**b**a]) with a tier-adjacent stem consonant. She argues that sibilant harmony is more phonetically motivated than voicing harmony because it has an articulatory advantage voicing harmony lacks: anteriority agreement reduces the number of distinct gestures the speaker needs to make while voicing agreement does not, since the glottal position for [–voice] cannot be maintained through the vowel intervening between two agreeing (voiceless) consonants. The two patterns are of equal complexity, requiring agreement between two instances of the same feature. Shapp's participants learned both types of alternation, but sibilant harmony was learned better than voicing harmony, supporting an effect of substantive bias rooted in ease of articulation.

Finley (2012) tested for substantive bias in the learning of rounding/backness harmony (hereafter simply rounding harmony) triggered by different types of vowels. Cross-linguistically, mid vowels are more likely to trigger rounding harmony than high vowels. This asymmetry is phonetically motivated: high round vowels sound rounder than mid round vowels, so rounding harmony improves the perceptibility of the round feature for mid vowels more than it does for high vowels (Terbeek 1977, Kaun 2004). In other words, mid vowels are better triggers of rounding harmony than high vowels because they derive a greater perceptual benefit from harmony. Finley exposed participants to a vowel harmony pattern in which stem vowels triggered rounding agreement on the suffix vowel (e.g. [kini-mi], [tudu-mu]; note however that participants heard only whole words and were given no indication of any morphological parse). In one condition, the stem vowels were mid while in the other they were high. Finley found that participants in the mid vowel condition learned the rounding harmony pattern better, as evidenced by their ability to apply it to novel forms. Participants in the high condition did not in fact learn rounding harmony; their performance was not significantly better than that of control participants. Thus learners favored the more perceptually natural harmony pattern.

Saltation is an unnatural and cross-linguistically rare pattern in which dissimilar sounds alternate while relatively more similar sounds do not. For example, if in a hypothetical language /p/ surfaced as [v] intervocalically, “jumping” over [b] or [f], but /b/ surfaced as [b] and /f/ as [f] in the same context, this would be unnatural because [p] alternates with [v] while the more similar [b] (in voicing) and [f] (in continuancy) do not. In his Experiment 1, White (2013) trained participants on an alternation that was potentially consistent with saltation and found that they extended the alternation to novel segments in such a way that the phonological system was not saltatory. That is, learners were biased toward the more natural, non-saltatory pattern. In Experiment 2, he showed that participants underlearned a saltatory alternation when explicitly trained on it, providing further evidence for a naturalness bias. Finally, he conducted an experiment with 12-month-old infants that also found a bias against saltatory alternations: when taught that a sound alternated with a relatively dissimilar sound (e.g. [p] with [v]), infants assumed that a more similar sound also alternated with that sound (e.g. [b] with [v]). In all three experiments, White established the anti-saltatory bias by comparing performance in critical conditions with performance in control conditions. For instance, in Experiment 1, participants in the control condition extended the alternation they had been taught much less to novel segments because they did not need to extend to novel segments to render the pattern they had learned non-saltatory.

White's study resembles Skoruppa, Lambrechts, & Peperkamp 2011 in that, in seeking to examine the effect of phonetic distance/similarity, both test the learning of alternations that differ in the number of features that change. As with the earlier study, then, White's finding that learners have trouble with saltation could be an effect of complexity bias rather than substantive bias. In Experiment 1, if learners extended the trained alternation to novel segments such that the system was not saltatory, they could use a simpler rule than if they applied only the trained alternation, maintaining a saltatory system. White acknowledges that a complexity bias account is possible, but he argues that it could not capture all the effects in his results. Experiment 1 taught participants that /p/ surfaced as [v] intervocalically (a two-feature change) and then tested whether participants thought /b/ and /f/ also surfaced as [v] (each a one-feature change) or surfaced faithfully. To make the system maximally non-saltatory and simple, participants should have extended the alternation to both /b/ and /f/. They did so, but not to the same degree: they alternated [b] with [v] more often than they alternated [f] with [v]. Complexity bias predicts no such difference because in each case one feature changes. Using confusion matrix data (Wang & Bilger 1973), however, White shows that [b] and [v] are more perceptually similar than [f] and [v], which could explain why participants endorsed a [b]~[v] alternation more than an [f]~[v] alternation. Thus there must still be a role for perceptual similarity among the factors that constrain phonological learning.

An interesting natural language study that investigates substantive bias effects over the course of language acquisition is van de Vijver & Baer-Henney 2014. German, like Turkish, has final devoicing, and so some, but not all, stem-final voiceless consonants "voice intervocalically" when the plural suffix is appended to a singular stem (e.g. [bɛɐk] ‘mountain’ to [bɛɐg-ə] ‘mountains’). Additionally, some, but not all, singular stems undergo fronting of their final back vowel in the plural (e.g. [kuː] ‘cow’ to [kyː-ə] ‘cows’). Intervocalic voicing is phonetically natural, promoting ease of articulation, while the vowel fronting process is not phonetically motivated. The rates at which German stems undergo each process from the singular to the plural are comparable. van de Vijver & Baer-Henney conducted a wug test in which 5-year-old, 7-year-old, and adult German speakers produced the plural forms of nonce singular nouns. They found that 5-year-olds overapplied natural intervocalic voicing and underapplied unnatural vowel fronting relative to their occurrence in the lexicon. Application of intervocalic voicing decreased with age while application of vowel fronting increased with age. The authors’ interpretation is that learners first lean on substantive bias but do so progressively less as their lexicon grows and becomes more reliable. With enough experience, German speakers approach frequency matching rather than privileging the more natural plural alternation. This study uncovers evidence for substantive bias while also demonstrating how its effects are modulated by age.

1.2.2.5 Substantive Bias Effects with Problematic Elements

Other studies have found effects of substantive bias, or at least effects partially consistent with substantive bias, but exhibit peculiarities that invite caution in taking the results as strong evidence for substantive bias. Wilson’s (2006) well-known AGL study takes as its starting point the claim that if two sounds alternate in a context in which they are less perceptually similar, they should also alternate in a context in which they are more perceptually similar (Steriade 2001b). This appears to be borne out in the typology of velar palatalization: languages that palatalize velar stops before backer vowels, where velar stops and palato-alveolar affricates are more distinctive, also palatalize them before fronter vowels, where velar stops and palato-alveolar affricates are less distinctive.

Wilson exposed participants to velar palatalization (/k/ → [t͡ʃ], /g/ → [d͡ʒ]) either before [i] or [e] and then tested whether they would generalize palatalization to the other context. Velar stops and palato-alveolar affricates are less perceptually similar before [e] than before [i], so the prediction from naturalness is that participants should extend palatalization from before [e] to before [i] but not vice versa.[6] Indeed, participants exposed to palatalization before [e] did generalize to the [i] context while those exposed to palatalization before [i] did not generalize to the [e] context.

This result nicely supports substantive bias, but the broader results introduce some inconsistencies. Participants were also taught that velar stops did not palatalize before [ɑ], and participants exposed to palatalization before [i] did not generalize to the [ɑ] context, which is as expected. However, participants trained to palatalize before [e] extended not only to the [i] context, as substantive bias predicts, but also to the [ɑ] context, which substantive bias does not predict, as velar stops and palate-alveolar affricates are more dissimilar before [ɑ] than before [e].[7] Another concern is that Wilson's participants did not apply the velar palatalization alternation they were taught at above chance rates in the test phase, suggesting they had not fully learned the pattern they were trained on. This is likely because the test phase was a production task, a more difficult AGL paradigm (cf. Peperkamp, Skoruppa, & Dupoux 2006). I do not

---

[6] Note that [i] is a high front vowel and [e] is a mid front vowel, but [i] is phonetically fronter than [e].

[7] Wilson's (2006:960) particular implementation of a grammatical model incorporating substantive bias actually predicts this extension. In his model, when the constraint that militates against palatalizing the velar stop is weakened sufficiently to allow palatalization to occur before [e], it is also weakened enough to allow palatalization before both [i] and [ɑ]. Thus Wilson considers the extension to the [ɑ] context to be further proof of the model's accuracy, but he acknowledges that it is an idiosyncrasy of his implementation of the substantive bias.

believe this makes Wilson's result uninterpretable: participants in both conditions underlearned the alternation relative to the input, but only in one condition did participants generalize the (underlearned) alternation to new contexts. That said, it is preferable to be able to demonstrate learning of the trained pattern before interpreting performance on untrained patterns.

Kimper (2016) takes up the same asymmetry in the triggers of rounding harmony that inspired Finley 2012 but from an implicational perspective. Because mid vowels benefit more, perceptually, from rounding harmony than high vowels, a language in which high vowels trigger rounding harmony should also exhibit mid vowels as triggers of rounding harmony; not to do so would make for an unnatural system. Mid vowel triggers of rounding harmony do not necessarily entail high vowel triggers, though. Kimper taught participants either high vowel-triggered or mid vowel-triggered rounding harmony using harmonizing suffixes and then tested whether participants extended rounding harmony to the other type of trigger. Though the test phase was not a production task (participants had to say whether plural forms they heard were acceptable in the language they had been trained on or not), only about half of his participants (the learners) acquired the trained harmony pattern well enough to be included in the final analysis. The results were consistent with substantive bias: high trigger condition participants treated mid vowels as rounding harmony triggers as often as they considered high vowels triggers while mid trigger condition participants treated high vowels as rounding harmony triggers less often than they considered mid vowels triggers.

Upon further examination, however, this asymmetric behavior only emerged among the highest-performing learners; among the lower-performing learners (who still performed above chance) there was no difference between the high trigger and mid trigger conditions. Additionally, the learners were able to describe the harmony pattern, suggesting they engaged in fairly explicit learning rather than the implicit learning associated with native language acquisition and thought to foster the emergence of substantive bias (Moreton & Pertsova 2016:288-289). Other properties of the experiment that point to explicit learning include the inclusion of a verification phase with feedback before the test phase and the division of the participants into learners (some of whom performed very well) and non-learners (Moreton & Pertsova 2016). Kimper's study seems to uncover an effect of substantive bias that nicely complements Finley's (2012) rounding harmony finding, but the effect is only present among a subset of his participants and comes out, surprisingly, as Kimper acknowledges, in the context of relatively explicit learning.

The study in Finley 2012 actually included an extension component like Kimper 2016 which is only reported in Finley 2008 (Experiment 8). In addition to testing how well participants had learned rounding harmony triggered by high vowels or mid vowels, she tested whether participants extended rounding harmony to triggers of the other type. For novel stems, mid trigger condition participants applied rounding harmony significantly more often when the trigger was a mid vowel than it was a high vowel, showing an unwillingness to extend to high vowel triggers. High trigger condition participants did not exhibit this difference, applying harmony as often when the trigger was a mid vowel than when it was a high vowel. This does not demonstrate the expected extension to mid vowel triggers, though, because as mentioned in the discussion of Finley 2012, the high trigger condition participants did not learn the pattern they were trained on. They may have simply been performing at chance in the two-alternative forced choice task for both trained and novel trigger types. In short, the behavior of the mid trigger condition participants is consistent with the substantive bias Kimper (2016) found while that of the high trigger condition participants is largely uninterpretable.[8]

Finley & Badecker's (2009) Experiment 1 also exhibits some similarity to Finley 2012 and Kimper 2016 in that in one condition participants were taught rounding/backness harmony triggered by high vowels and tested on whether they extended it to mid vowel triggers. Participants did in fact extend from high triggers to mid triggers, consistent with one half of Kimper's result. (In this condition, participants were also taught harmony triggered by low vowels, but they failed to learn it, perhaps because low vowels harmonized only in backness and not in rounding.) Since Finley & Badecker (2009) were investigating different questions, there was no complementary condition that tested whether participants extended rounding harmony from mid vowel triggers to high vowel triggers, but their experiment partially corroborates Kimper's findings.

Finley & Badecker's (2012) Experiment 1 is very similar to Finley 2012, but it investigates the targets of height harmony instead of the triggers of rounding harmony. Typologically, front vowels are favored over back vowels as targets of height harmony. Finley & Badecker entertain several possible phonetic motivations for this asymmetry. As mentioned above, high round vowels are perceived as rounder than mid round vowels. Height harmony is often construed as vowel lowering, and lowering [u] to [o] makes it harder to hear that it is round. Height harmony thus hurts the perceptibility of back (round) vowels, so it is better to

---

[8] The reason I consider this case uninterpretable while arguing that Wilson's (2006) result remains interpretable despite participants not learning to apply velar palatalization more than 50% of the time has to do with the type of task. Wilson's experiment was a production task. By spontaneously producing even a low proportion of palatalized forms, his participants showed they had learned to apply an alternation, i.e. they showed an effect of training. Finley's task, on the other hand, was a two-alternative forced choice between a harmonic form and a disharmonic form. The high trigger condition participants' willingness to apply rounding harmony to novel mid vowel triggers as much as to trained high vowel triggers cannot be taken as extension because they chose harmonic forms at chance levels in both cases, a result that could have been achieved by guessing on all trials.

avoid them as targets. Alternatively, [i] and [e] share the same constriction location while [u] and [o] have different constriction locations, so height harmony is articulatorily more difficult for back (round) vowels than for front vowels. Finally, the phonetic precursor to height harmony, F1 coarticulation, may be stronger for front vowels than for back vowels, making height harmony more phonetically grounded for front vowels (cf. Pycha et al. 2003, Moreton 2008).

Finley & Badecker trained participants on height harmony targeting either front vowels (e.g. [getog-**e**], [dunig-**i**]) or back vowels (e.g. [getog-**o**], [dunig-**u**]) (both front and back vowels triggered height harmony) and tested their learning of the pattern and their extension of height harmony to the other type of target. Participants in the front target condition learned height harmony but did not extend it to back vowel targets while participants in the back target condition did not learn to apply height harmony to back vowels but nevertheless applied it to front vowel targets. This result is consistent with a substantive bias toward front vowel targets of height harmony, but it is surprising that back target condition participants applied height harmony to the new type of target when they were not able to apply it at above chance levels to the type of target they had been trained on. It does not appear to be the case that the participants came to the experiment with an a priori inclination to apply height harmony to front vowel targets because control participants did not do so. The authors speculate that the back target condition participants may have learned to harmonize front vowels for height from hearing harmonic stems containing front vowels in training, but front target condition participants heard harmonic stems containing back vowels in training and did not extend height harmony to back vowels in test.

Baer-Henney's (2015) third study (Chapter 4) is unusual in testing for a gradient effect of substantive bias. She examines intervocalic voicing, an articulatorily natural process, but she distinguishes degrees of naturalness within intervocalic voicing depending on place of articulation. The larger the cavity between the stop closure and the glottis, the easier it is to maintain voicing of a stop (Westbury & Keating 1986), so voicing labial stops is more natural than voicing coronal stops, which is in turn more natural than voicing dorsal stops.

Baer-Henney trained German speakers on intervocalic voicing[9] (e.g. /fina **p**unə/ → [fina **b**unə]) of labials, coronals, or dorsals and tested how well they learned the alternation, as well as how much they extended it to untrained places of articulation. Her participants learned intervocalic voicing for dorsals, the least natural case, worse than for labials or for coronals, suggesting an effect of substantive bias. Additionally, participants trained on intervocalic voicing of labials and coronals extended the process less to dorsals than to the other place of articulation they had not been trained on, also in line with the predictions of substantive bias. However, the overall pattern of extension results does not seem consistent with substantive bias. Given the gradient naturalness of intervocalic voicing at different places of articulation, we expect implicationals: intervocalic voicing at a given place should entail intervocalic voicing at more anterior places but not at more posterior places. In terms of Baer-Henney's experiment, this would translate to least extension of intervocalic voicing from labials to other places and most extension from dorsals to other places. Instead, participants trained on intervocalic voicing of labials seemed to extend the most to other places of articulation, more than participants trained on the voicing of coronals or dorsals. Like Wilson 2006, Baer-Henney's study yields some

---

[9] From the point of view of its global phonological grammar, German does not have intervocalic voicing; instead, it has final devoicing. Final devoicing can give rise to the appearance of intervocalic voicing for some stems in, for instance, plural formation (see van de Vijver & Baer-Henney 2014, which tests German speakers' acquisition of this native "intervocalic voicing"), so it is probably incorrect to say that German speakers have no experience with the artificial pattern being taught here.

evidence for substantive bias but also exhibits effects that do not fit in with a substantive bias account.

The studies reviewed in this section appear to find genuine effects of substantive bias, but they also suffer from problems that make their support for substantive bias less clear-cut. Some of them show patterns of results that are not entirely consistent with substantive bias (Wilson 2006, Baer-Henney 2015) or are simply unusual (Finley & Badecker 2012, Kimper 2016). A lack of learning in one condition renders the results of Finley's (2008) Experiment 8 partially uninterpretable, and Kimper's (2016) substantive bias effect only holds for a specific subset of participants. Experiments testing for substantive bias often yield results that are challenging to interpret and weaken the study's overall support for naturalness bias.

1.2.2.6 Null Results

I now consider alternation learning studies that have failed to support substantive bias by finding equal learning of phonological patterns differing in their naturalness. Pycha et al. (2003) taught participants artificial languages that exhibited vowel backness harmony and vowel backness disharmony, as manifested by the allomorphs of a plural suffix. As mentioned above, they considered vowel harmony natural, because the phonological pattern can develop from phonetic precursors in the speech signal, but vowel disharmony unnatural. While participants in the vowel harmony condition were more accurate at determining whether novel singular-plural pairs conformed to the pattern they had been trained on than participants in the vowel disharmony condition (86% correct vs. 75% correct), performance in the natural harmony condition was not significantly better than in the unnatural disharmony condition. There was a tendency for the natural pattern to be learned better, but the experiment does not provide strong support for substantive bias.

In a sister study to Peperkamp, Skoruppa, & Dupoux 2006, Peperkamp & Dupoux (2007) taught French speakers the same patterns as in the earlier study, namely, natural allophonic distributions of voiced and voiceless obstruents, demonstrated by intervocalic voicing alternations, and unnatural allophonic distributions relating arbitrary consonants (e.g. [f] is the intervocalic allophone of /ɡ/). Instead of a production task, though, they tested participants with a word-picture matching task that probed learning of the allophonic distributions. The natural and unnatural languages were apparently learned equally well, though no statistical analysis directly compared learning of the two patterns. This result is surprising since, as I pointed out in discussing Peperkamp, Skoruppa, & Dupoux 2006, the unnatural patterns are more complex than the natural ones, and we might expect complexity bias, if not substantive bias, to favor the natural languages. The task in Peperkamp & Dupoux 2007 was easier, though. Moreover, Peperkamp & Dupoux believe their participants learned all the distributions segment-by-segment, since those learning the natural patterns did not generalize to the whole natural class of obstruents in test when exposed to most of it in training.

Finley & Badecker’s (2009) Experiment 3 resembles Finley 2012 in that it tests the learning of rounding harmony, but whereas Finley 2012 examined different triggers of rounding harmony, Finley & Badecker examined different targets. Cross-linguistically, high vowels are favored as targets of rounding harmony: if high vowels are targets, that does not necessarily mean other vowels will be, but if mid vowels are targets, high vowels should be too (Kaun

2004).[10] The roundedness of high rounded vowels is more perceptible than that of mid rounded vowels (Terbeek 1977, Kaun 2004), so conceivably high targets of rounding harmony are more useful for signaling the roundedness of the triggering vowel and are also simply more likely to be perceived as participating in rounding harmony. Finley & Badecker (2009) taught participants a vowel harmony pattern in which both high and mid stem vowels triggered rounding harmony on the suffix vowel. In one condition, participants were only exposed to high vowel targets of rounding harmony (e.g. [bide-m**i**], [podo-m**u**]), and in the other they were only exposed to mid vowel targets of rounding harmony (e.g. [bide-g**e**], [podo-g**o**]). Participants were then tested on their ability to choose harmonic forms with suffixes with high *and* mid vowels. An effect of substantive bias would emerge as greater extension of rounding harmony from mid vowel targets to high vowel targets than from high vowel targets to mid vowel targets. Instead, Finley & Badecker found equal extension from one type of target to the other.

Finley's (2008) Experiment 7, a follow-up to the experiment reported in Finley & Badecker 2009, made stem vowels identical (e.g. [duku], [tepe]) so that they would agree in height. This was to prevent participants from learning that vowels of the other height (either high or mid) could be targets of rounding harmony based on evidence from the stems. The follow-up experiment yielded the same result: equal extension of rounding harmony from one target height to the other. Although there seem to be plausible perceptual reasons for high vowels to be more natural targets of rounding harmony than mid vowels, Kaun (2004) argues for a diachronic, perception-driven explanation of the favoring of high vowel targets. If she is right, this may

---

[10] In some rounding harmony systems, only non-high vowels participate in rounding harmony; in such systems mid vowels can be targets while high vowels are not (Kaun 2004).

explain the lack of a naturalness bias effect in these experiments (see also Finley 2008 for further discussion of this null result).

Do, Zsiga, & Havenhill (2016) compared the learnability of postnasal voicing and postnasal devoicing. For reasons relating to nasal leakage and the timing of the raising of the velum, postnasal voicing is considered articulatorily natural, with postnasal devoicing its unnatural counterpart (Hayes & Stivers 1995, Pater 1999). Do, Zsiga, & Havenhill taught the artificial patterns with a VN- plural prefix that triggered voicing alternations on the following stem-initial consonant (e.g. singular [**p**abli] and plural [am-**b**abli]). They also tried to foster more or less implicit learning of the voicing alternations: the plural prefix also exhibited vowel harmony, and in Experiment 1 participants were instructed to learn how plurals were formed (explicit learning of the voicing alternation) while in Experiment 2 they were only instructed to learn when each prefix vowel was used (implicit learning of the voicing alternation). The authors also manipulated the proportion of training items that exhibited the voicing alternation (cf. Baer-Henney & van de Vijver 2012). The test phase consisted of a forced choice task in which participants had to choose the correct plural form for novel words. In both Experiment 1 and Experiment 2, participants learned postnasal voicing and devoicing equally well, showing no preference for the more natural process.

Lysvik (2018) tested the learnability of final stop devoicing, a phonetically natural alternation, and final stop voicing, a phonetically unnatural alternation. His participants, native Norwegian speakers, learned to apply both processes to an equal degree, yielding no evidence for substantive bias. I discuss this study in more detail in Chapter 4.

The AGL experiments reviewed here all found equal learning of phonetically natural and unnatural alternation patterns (and even, in the case of Peperkamp & Dupoux 2007, equal learning of a simple, natural pattern and a complex, unnatural pattern). While a null result does not constitute evidence that substantive bias does not play a role in phonological learning, an accumulation of null results chips away at the plausibility of the substantive bias hypothesis. That is, if in experiment after experiment participants acquire phonetically unmotivated or unnatural patterns as well as they acquire formally equivalent phonetically motivated patterns, it begins to look like there is no synchronic bias that favors the learning of natural patterns. As I showed in Section 1.2.2.3, many studies that claim to have uncovered evidence for substantive bias may only have found evidence for complexity bias, so the support for substantive bias in the literature is sparser than sometimes reported. Moreton & Pater (2012b) cite additional substantive bias studies that turned up null results, and their conclusion is that there is scant evidence for this type of learning bias.

While it is true that null results are numerous (and perhaps underreported) in the substantive bias literature, it is important to consider explanations for these null results other than the non-existence of substantive bias. Finley (2012) notes that one reason a study might not find a difference in learnability between a natural and an unnatural pattern is that testing occurred too late. Participants may learn an unnatural pattern more slowly but ultimately be able to learn it as well as the natural pattern. If this is the case, then in order for a difference to be detected testing must happen before learning of the unnatural pattern has caught up with that of the natural pattern. This type of learning trajectory, with initial differences that later faded, has been observed before (Schane, Tranel, & Lane 1974). Careful timing or different methods of testing learning may be necessary to detect subtle effects of substantive bias.

Another cause of null substantive bias results may be the failure of many experiments to tap into linguistic learning. If AGL experiments tend to elicit more explicit or domain-general pattern learning on the part of participants, they may be unable to detect effects of language-specific naturalness bias which nevertheless exist. I discussed the issue of what type of learning AGL is able to probe in Section 1.1.2.

A third possible reason for the prevalence of null results in studies testing for substantive bias is that only some types of substantive bias exist. The term covers any kind of bias toward phonetically natural patterns, but the phonetic advantage that renders a pattern more natural may be based in perception or in articulation. If, for instance, only perceptual naturalness biased phonological learning, then experiments testing for articulatorily-rooted substantive bias would always turn up null results. This would give the appearance of weak support overall for substantive bias, when in fact splitting substantive bias into subtypes would reveal a pattern. This is precisely the argument I will advance in Chapter 5.

1.3 Substantive Bias in Phonotactic Learning

Investigations of substantive bias have focused mostly on the learning of alternations. Fewer studies have tested for substantive bias in the learning of phonotactics. In this section, I review the existing literature on substantive bias in phonotactic learning, as this is the focus of the experiments in this dissertation.

1.3.1 AGL Experiments

Seidl & Buckley 2005 exposed 9-month-old infants to phonetically grounded and arbitrary phonotactic patterns and tested whether the infants learned the phonetically grounded patterns better. In Experiment 1, the grounded pattern featured fricatives and affricates only intervocalically, with no stops in intervocalic position (e.g. [pasat], [bazu]), while the arbitrary pattern featured fricatives and affricates only word-initially, with no stops in that position (e.g. [sapat], [zabu]). In Experiment 2, in the grounded pattern initial labial consonants were always followed by round vowels and initial coronal consonants by front vowels (e.g. [**po**va], [**si**du]) while in the arbitrary pattern initial labial consonants were followed by high vowels and initial coronal consonants by mid vowels (e.g. [**pi**va], [**so**du]). In both experiments, infants learned the grounded and arbitrary patterns equally well: they listened longer to test words that did not conform to the phonotactic pattern they had been trained on, showing an ability to distinguish conforming and nonconforming words. Thus Seidl & Buckley found no support for substantive bias in infants' phonotactic learning.

Another infant phonotactic learning study that may bear on the substantive bias question is Saffran & Thiessen 2003. In their Experiment 2, 9-month-old infants were exposed to words that featured either voiceless stops in onsets and voiced stops in codas (e.g. [todkad]) or voiced stops in onsets and voiceless stops in codas (e.g. [dakdot]). As Seidl & Buckley (2005) point out, the latter pattern is relatively more phonetically grounded than the former, but infants learned both patterns to the same degree, exhibiting no effect of substantive bias. Saffran & Thiessen's Experiment 3 repeated Experiment 2 except with arbitrary (i.e. not corresponding to a natural class) restrictions on which stops could occur in onset vs. coda. Infants were unable to learn this pattern, which, as Moreton & Pater (2012a) note, supports complexity bias.[11]

Turning to studies with adults, Skoruppa & Peperkamp (2011) taught French speakers artificial variants of French that featured vowel harmony (Harmonic French) or vowel

---

[11] These results should be interpreted with caution in light of Cristia's (2018) meta-analysis of infant phonotactic learning studies, which found an average effect size of zero (see also Footnote 2). Her finding casts doubt on whether infants actually demonstrate any phonotactic learning in AGL experiments. Seidl & Buckley 2005 was one of the sixteen studies included in her meta-analysis.

disharmony (Disharmonic French). Both the harmony and disharmony patterns were static generalizations; there were no alternations. While vowel harmony is typologically common and phonetically motivated and vowel disharmony is less common and not obviously phonetically motivated, participants learned Harmonic French and Disharmonic French equally well, as evidenced by their ability to identify words that could belong to the language they had been exposed to. The study thus yielded no evidence for substantive bias. Skoruppa & Peperkamp did find a complexity bias: participants performed worse on a variant of French featuring a mixed pattern of harmony and disharmony whose expression in features was more complex than the featural expressions of the generalizations in Harmonic and Disharmonic French.

Myers & Padgett (2014) conducted a series of experiments examining generalization of final devoicing from the domain of the phrase to the domain of the word. In Experiment 1, they taught participants either a natural phonotactic restriction against phrase-final voiced obstruents or an unnatural phonotactic restriction against phrase-final voiceless obstruents. Both phonotactic systems were learned to the same degree, so the experiment provided no support for substantive bias.[12]

---

[12] Myers & Padgett (2014) actually claim that “final devoicing” (i.e. the static pattern in which final voiced obstruents were disallowed) was learned better than “final voicing” (i.e. the static pattern in which final voiceless obstruents were disallowed). Experiment 1 tested participants not only on phrase-final obstruents but on word-final obstruents (since the goal of the experiment was to test whether a phrase-final restriction would be generalized to the word-final domain). The phrase-final “devoicing” pattern was generalized somewhat to word-final position while the phrase-final “voicing” pattern was not, suggesting a difference in the degree of domain generalizability of the two patterns. However, Myers & Padgett’s claim that “final devoicing” was learned better comes not just from the difference in willingness to generalize to the word-final domain but from an analysis of the combined phrase-final and word-final test trials that yields a significant main effect of condition: participants in the devoicing condition had a higher correct response rate (accepting final voiceless obstruents and rejecting final voiced obstruents) than participants in the voicing condition. Despite the lack of a significant interaction of condition and domain (phrase-final vs. word-final), this effect looks as though it is due to the poorer performance of participants in the voicing condition on word-final test trials (i.e. the lack of generalization from the phrase-final to word-final domain in the voicing condition). There is no analysis of just the phrase-final test trials that shows there is a significant difference in how well the two phrase-final training patterns were learned. In other words, Myers & Padgett do not demonstrate that the phrase-final “devoicing” pattern was better learned than the phrase-final “voicing” pattern.

In a similar experiment, Greenwood (2016) compared the learnability of a language with a natural phonotactic restriction against word-final voiced obstruents and a language with an unnatural phonotactic restriction against word-final voiceless obstruents. Greenwood conducted two versions of her experiment, one using careful speech, meaning that the stimuli were hyperarticulated, and one using casual speech, meaning that the stimuli were produced naturally. She expected better learning of the natural pattern only in the casual speech condition and equal learning of the two languages in the careful speech condition.[13] Instead, in the casual speech condition, both languages were learned to the same degree. In the careful speech condition, the unnatural language that disallowed word-final voiceless obstruents was learned better than the natural language that disallowed word-final voiced obstruents. This pattern is the opposite of the one that would have constituted evidence for substantive bias. None of the results of the voicing experiment support the idea that natural phonotactic generalizations are learned better than unnatural ones.

Investigations of substantive bias in phonotactics have also looked beyond segmental patterns to patterns involving suprasegmental features. Carpenter (2006) investigated the learning of natural and unnatural stress systems by English and French speakers. In the first set of experiments, low vowels attracted stress in the natural system while high vowels attracted

---

[13] Greenwood's explanation for why the restriction against word-final voiceless obstruents is unnatural is based on Universal Grammar and the universal constraint set, but the idea that voiceless obstruents are articulatorily natural and voiced obstruents articulatorily unnatural in word-edge position (Westbury & Keating 1986) seems to be implicit in her discussion. Since the experiment does not use a production task, participants must rely on their knowledge of articulatory difficulty (as opposed to in-experiment articulatory experience; see Footnote 14). It may seem odd to test careful speech and casual speech conditions when the type of naturalness bias being tested is rooted in articulation, not perception, but participants must accurately perceive the experimental stimuli in order to deploy their phonetic knowledge about ease of articulation. (Of course, Greenwood hypothesizes that this knowledge will *not* bias them in the careful speech condition.) In the casual speech condition, final voiced obstruents may be partially devoiced, leading participants to conclude that there is no restriction on the voicing of final obstruents to learn. This would impede learning of the unnatural phonotactic generalization.

stress in the unnatural system. Low vowels are inherently longer than high vowels (Lehiste 1970). They are thus intrinsically prominent, and this intrinsic prominence, combined with the fact that greater duration is one of the acoustic correlates of stress, makes low vowels better bearers of stress (Kenstowicz 1997). Both English and French speakers learned the natural stress system better than the unnatural one, though the English speakers demonstrated better learning overall, probably due to their L1 experience with a system in which stress is not fixed.

In the second set of experiments, heavy syllables attracted stress in the natural system while light syllables attracted stress in the unnatural system. The natural stress system exhibits the well-attested Weight-to-Stress principle (Prince 1990); again, stressing syllables that are already prominent in some way makes them more perceptually distinguishable from unstressed syllables (see Greenwood 2016). In this case, though, both English and French speakers learned the natural and unnatural systems equally well. Carpenter found an effect of naturalness in the learning of stress systems conditioned by vowel height but not in the learning of stress systems conditioned by syllable weight. She speculates that the unnatural system in which light syllables attracted stress was learned as well as its natural counterpart because, due to the way her stimuli were created, the stressed light syllables had more perceptual energy than the stressed heavy syllables. This likely made them quite perceptible and perhaps even caused them to be perceived as heavy syllables too, in which case the natural and unnatural systems did not actually differ.

Greenwood (2016) also tested the relative learnability of weight-sensitive stress systems that varied in naturalness and complexity. Two systems were like Carpenter's, a natural system in which heavy syllables attracted stress and an unnatural system in which light syllables attracted stress. In the third system, heavy syllables containing non-high vowels attracted stress; this pattern is natural (cf. Carpenter's (2006) vowel height-based systems) but relatively complex and also unattested. Of the three systems, participants only learned the natural, and simple, system in which heavy syllables attracted stress. Participants' failure to learn the complex natural system is another complexity bias result. As for the unnatural system, Greenwood concludes that participants failed to learn it because stress was perceived less well in this condition. A perception test showed that stress was perceived much less accurately in the unnatural system (specifically in words in which a light syllable attracted stress away from its default initial position).

Another of Greenwood's (2016) studies compared learning of a natural phonotactic pattern in which only sonorants were legal codas and an unnatural phonotactic pattern in which only voiceless stops were legal codas. She argues that sonorants make more natural codas than voiceless stops because they are less susceptible to the perceptual weakening that occurs in coda position. As in her obstruent voicing experiment, Greenwood conducted careful speech and casual speech versions of her experiment. In the careful speech condition, the natural and unnatural patterns were learned equally well, but in the casual speech condition, the natural pattern was learned better than the unnatural one. A perception test showed that voiceless stop codas in the casual speech condition were often not perceived (i.e. their presence was not heard). This would explain why participants trained on the unnatural coda sonority phonotactic were worse at choosing pattern-conforming items in the casual speech condition. Again, Greenwood concludes that a naturalness bias effect only comes out when the unnatural pattern is less perceptible in the experiment itself.

Greenwood's position is worth unpacking. Her claim is that naturalness bias effects only emerge in AGL experiments when the phonetic (dis)advantage that distinguishes the natural and unnatural patterns is reproduced in the experimental stimuli. In that case, though, the difference in learning cannot be attributed to a substantive bias independent of the perceptibility of the stimuli used to expose participants to each pattern. One expects a less perceptible pattern to be learned worse than a more perceptible pattern in a laboratory setting. The implicit assumption in the AGL literature on substantive bias is that the bias stems from learners' *knowledge* of phonetic naturalness, whether innate or acquired, and that it will manifest even under perfect conditions with maximally perceptible stimuli.[14] Indeed, this is the true test of substantive bias. Greenwood contends that experiments testing for substantive bias have not always adequately controlled the relative perceptibility of the patterns being compared and that those that uncover effects of substantive bias have done so because the phonetic disadvantage of the unnatural pattern was inadvertently reproduced in the exposure stimuli. That is, apparent substantive bias effects are only due to what might be called in-experiment channel bias. She conducts experiments with careful speech (eliminating the phonetic disadvantage) and casual speech (exhibiting the phonetic disadvantage) conditions to make her argument, and the results of her coda sonority experiment support her claim: when a bias against the unnatural pattern could only have stemmed from the knowledge participants brought to the experiment, no effect emerged, but when it could have stemmed from participants misperceiving the experimental stimuli, an effect did emerge.

Greenwood's work offers a caution for AGL studies: to avoid a confound, experimental stimuli must not reproduce the perceptual (dis)advantage that underpins the argument for one pattern being more natural than another. Most studies seem to take pains to ensure that the

---

[14] Note that this only applies to phonetic naturalness that is rooted in perception. For phonetic naturalness rooted in articulation, phonetic knowledge cannot be separated from the physical reality of the phonetic (dis)advantage if the experimental task involves production. For instance, in Baer-Henney's (2015) intervocalic voicing study, participants had to intervocalically voice labials, coronals, and velars and so could not escape the greater articulatory difficulty of voicing more posterior obstruents. An experiment can test for a substantive bias based only on learners' knowledge of articulatory difficulty if it does not use a production task.

stimuli are produced clearly, but sometimes the perceptual (dis)advantage may be subtler and/or less under the control of the (usually phonetically-trained) speaker producing the stimuli. For example, in Finley 2012, the reason mid vowels are more natural triggers of rounding harmony than high vowels is because rounded mid vowels are perceived as less rounded than rounded high vowels and therefore triggering harmony to cue roundedness benefits mid vowels more. This study found an effect of substantive bias. To show that this effect was due to participants' phonetic knowledge rather than the properties of the stimuli, one would have to demonstrate that rounding in the mid vowels was perceived as well as rounding in the high vowels in the experiment. A rater deemed all of Finley's stimuli to have clear vowels, and participants completed a separate AXB perception task, in which they exhibited no difference in their ability to distinguish rounding/backness in English high and mid vowels. The experimental stimuli themselves were not subject to a separate perception task, though (see Finley 2012:1560 for discussion of the results of the AXB test and their relation to the bias toward mid vowel rounding harmony triggers uncovered in the experiment.) To rule out the possible confound Greenwood raises, it would be most prudent to separately test the perceptibility of all stimuli in an AGL experiment.

Though Greenwood's caveat is warranted, her claim that all apparent substantive bias effects in AGL studies can be attributed to in-experiment channel bias is perhaps too strong. One cannot dismiss a substantive bias result for this reason if a separate perception task has not shown that the experimental stimuli demonstrating the unnatural pattern were in fact degraded. Consider again Carpenter's (2006) first experiment on stress and vowel height. She found that a natural system in which low vowels attracted stress was learned better than an unnatural system in which high vowels attracted stress. Low vowels are more natural bearers of stress because they have inherently longer duration than high vowels, but Carpenter was careful to make the low vowels and high vowels in her experimental stimuli of equal duration, thus removing low vowels' phonetic advantage in the experiment itself. She did so precisely to avoid the confound Greenwood discusses (Carpenter 2006:73), and she still obtained an effect of naturalness bias. Greenwood suggests the position of stress may still have been perceived less well in Carpenter's unnatural system, but no perception test shows this. Given Carpenter's care in creating the stimuli, there is little reason to think her unnatural condition reproduced a perceptual disadvantage that hurt participants' performance. In short, Greenwood is right to advise caution, but it would be premature to conclude that no study has uncovered a true effect of substantive bias.

1.3.2 Natural Language Studies

Natural languages studies have also been used to test for substantive bias in phonotactic learning, and unlike nearly all the AGL studies reviewed above, this set of studies has purported to find a learning advantage for natural phonotactic patterns. Hayes & White (2013) used the UCLA Phonotactic Learner (Hayes & Wilson 2008) to uncover natural and unnatural phonotactic constraints on the English lexicon. The constraints had similar weights (i.e. they were of similar strength in the lexicon), but the unnatural ones had little typological or phonetic support. Hayes & White found that English speakers underlearn the unnatural phonotactic constraints, as evidenced by a blick test (a phonotactic well-formedness judgment task (Chong 2017, Scholes 1966)). Unlike English nonce words that violated the natural constraints, nonce words that violated these unnatural constraints were not judged as strongly ill-formed relative to similar control words. Thus English speakers had not internalized unnatural phonotactic constraints that were nevertheless supported by the lexicon, suggesting an effect of substantive bias.

It is possible that the effect that emerged in this study reflected complexity bias, however, because as Prickett (2018) has also noted, the unnatural constraints were more complex than the natural ones. The natural constraints related two instances of the same feature more often than the unnatural constraints did, and such intradimensional generalizations have been shown to be easier to learn than interdimensional generalizations (i.e. those relating instances of two different features) (Moreton 2008, 2012). Additionally, as Hayes & White themselves acknowledge and as Greenwood (2016) points out, the natural constraints were nearly all about dependencies between two consonants while the unnatural constraints mostly concerned dependencies between vowels and consonants. Such within-tier dependencies may be easier to learn than between-tier dependencies (cf. Moreton 2012). Finally, the unnatural constraints simply required more features to state than did the natural constraints. For all these reasons, the unnatural constraints may have been underlearned due to their complexity rather than due to their unnaturalness.

Prickett (2018) conducted a follow-up study with the goal of avoiding the confound of naturalness and complexity that cast doubt on the effect of naturalness found by Hayes & White. He modeled his experiment on theirs but manipulated both naturalness and complexity, which were fully crossed. The constraints he selected were categorized as Natural Simple, Natural Complex, Unnatural Simple, and Unnatural Complex. Prickett found effects of both naturalness and complexity: nonce words were rated worse than their control word counterparts when they violated natural constraints than when they violated natural constraints and worse when they violated simple constraints than when they violated complex constraints. There was also an interaction of naturalness and complexity whereby nonce words that violated Natural Simple constraints were judged even more ill-formed relative to their controls.

Prickett's naturalness effect is still not entirely reliable, however, because even within the simple and complex sets of constraints, the natural constraints were still simpler than the unnatural constraints. The Natural Simple constraints required 2 to 2.5 features to express, by Prickett's feature counting method, and always related two instances of the same feature while the Unnatural Simple constraints required 2.5 to 3 different features. The Natural Complex constraints required 3 to 4 features, with only one out of four constraints requiring 4, while the Unnatural Complex constraints required 3.5 to 4 features, with three out of four constraints requiring 4. Thus the naturalness effect in this study could still be driven at least in part by complexity, weakening the evidence for substantive bias.[15]

Another study modeled on Hayes & White 2013 is Prickett 2015, which also manipulated and crossed constraint naturalness and constraint complexity but additionally sought to compare naturalness and complexity effects on learners' L1 vs. L2. The participants were native speakers of Canadian French or Canadian English who spoke both French and English, and the study resembled Prickett 2018 but in both French and English versions, both of which all participants completed. The French version of the experiment suffered from design problems that made the results largely uninterpretable, but for English Prickett found effects of naturalness and of complexity, for both the native French speakers and the native English speakers. Moreover, in

---

[15] Within the simple and complex groups, the natural constraints also had on average slightly higher weights than the unnatural constraints. (This was determined by looking up the weights of Prickett's constraints, which were drawn from the English grammar Hayes & White obtained using the UCLA Phonotactic Learner, in the online supplementary materials for Hayes & White 2013: https://linguistics.ucla.edu/people/hayes/Phonological-Naturalness/Grammar.txt.) This means that the naturalness effect could have been due to the natural constraints being stronger in the grammar (i.e. having greater support in the English lexicon) rather than due to their being more phonetically natural. The difference in weights was small, though, particularly within the simple group.

this study, unlike in Hayes & White 2013 and Prickett 2018, the natural constraints were not obviously simpler than their unnatural counterparts.

This series of related studies claims that substantive bias exists by arguing that English speakers have better internalized natural phonotactic constraints on their language than unnatural phonotactic constraints. In most cases, though, the naturalness effect is unreliable because complexity was confounded with naturalness.

Supposing for a moment that Hayes & White 2013 and subsequent studies did uncover a true effect of substantive bias, it is worth asking why they did so when other phonotactic learning studies did not. Prickett (2015, 2018) observes that substantive bias effects seem to emerge only in native language studies, not in AGL experiments. Hayes & White 2013 and Prickett's studies used a nonce word rating task to test for biased learning of the phonotactics of speakers' own native language (or second language, in the case of the Canadian French speakers). In contrast, Myers & Padgett 2014, Greenwood 2016, and others tested participants' learning of unfamiliar phonotactic patterns they first had to be taught, and Skoruppa & Peperkamp 2011 also taught participants a non-native pattern, even if the artificial language was derived from French. It is possible that substantive bias is not detectable through artificial grammar learning but requires paradigms that probe speakers' knowledge of their own languages. However, it should be noted that at least for phonological learning in general, this does not always hold true, since there are AGL experiments involving alternation learning that have found effects of substantive bias (Finley 2012, White 2013). The differences between AGL experiments and natural language studies were discussed in more detail in Section 1.1.2. Ideally, agreement between both types of studies will lead to new insights into the nature of phonotactic learning.

Summing up, Saffran & Thiessen (2003), Seidl & Buckley (2005), Carpenter (2006), Skoruppa & Peperkamp (2011), Myers & Padgett (2014), and Greenwood (2016) all failed to find effects of substantive bias on phonotactic learning in AGL experiments. The only positive AGL result was Carpenter's finding that a natural stress system in which low vowels attracted stress was learned better than an unnatural system in which high vowels attracted stress. On the other hand, Hayes & White (2013) and Prickett (2015, 2018) argue, based on evidence from English blick tests, that naturalness does bias phonotactic learning, but most of their results may be reinterpreted as supporting a complexity bias. Skoruppa & Peperkamp (2011) also found an effect of complexity bias on phonotactic learning.

1.4 Motivation and Outline of the Dissertation

1.4.1 A New Approach to Testing for Substantive Bias in Phonotactic Learning

The research question that drives this dissertation is whether phonetic naturalness biases phonotactic learning. The great majority of the substantive bias literature probes alternation learning, so there remains something of a research gap. Of the phonotactic learning studies reviewed in Section 1.3, some do not adequately control for complexity in testing for an effect of naturalness (Hayes & White 2013, Prickett 2018). Carpenter (2006) found a naturalness bias effect in the learning of vowel height-sensitive stress systems, but other experiments yielded null results or apparent naturalness bias effects argued to stem from experiment-internal channel bias (Saffran & Thiessen 2003, Seidl & Buckley 2005, Skoruppa & Peperkamp 2011, Myers & Padgett 2014, Greenwood 2016). Skoruppa & Peperkamp 2011 pitted vowel harmony against vowel disharmony, but this may not be the best test of substantive bias if disharmony (and dissimilation more broadly) is also natural (see discussion in Ohala 1992, Pycha et al. 2003, Wilson 2003, and Baer-Henney 2015). In short, there is still room for exploring the role of phonetic naturalness in phonotactic learning.

Moreover, I take a different approach to testing for substantive bias in phonotactic learning by investigating not just phonotactic restrictions but phonotactic implicationals about the existence of contrasts in different positions. Given evidence for a contrast in one position, what do learners assume about that contrast's existence in other positions? I hypothesized that implicitly asking learners to compare the existence of contrasts across positions might cause an effect of substantive bias to emerge when simply testing the learnability of a specific phonotactic constraint does not. For example, both Myers & Padgett (2014) and Greenwood (2016) taught restrictions against voiced or voiceless obstruents in final positions, and participants were equally willing to learn both restrictions (or at least did not exhibit the expected preference for a restriction against voiced obstruents). In these experiments, participants simply had to learn a ban on a particular type of sound in a particular position.

In Chapter 3, I also teach phonotactic patterns of obstruent voicing, but I investigate a phonotactic implicational: a word-final obstruent voicing contrast entails a word-initial obstruent voicing contrast, but not vice versa. Due to the perceptibility of the obstruent voicing contrast, it is more natural for a language to exhibit this contrast only word-initially than only word-finally. I exposed participants to an obstruent voicing contrast in either word-initial or word-final position but withheld whether the contrast was also present in the other position (participants were exposed to obstruents in the other position, but not to a voicing contrast). In the test phase, participants demonstrated to what extent they thought the obstruent voicing contrast existed in the other position. In Myers & Padgett 2014 and Greenwood 2016, participants are invited to posit a constraint like *D# (where D is a voiced obstruent). By testing phonotactic implicationals about the distribution of contrasts, I invite participants to decide whether a contrast should exist in a given position given its existence in another position. This is what distinguishes the paradigm from earlier ones. While the artificial languages I teach can also be mastered with a specific phonotactic constraint (e.g. if participants are exposed to an obstruent voicing contrast word-initially and to voiceless obstruents word-finally, they may learn *D#, excluding the possibility of a word-final obstruent voicing contrast), the novel element of comparison across positions is also present. It is also worth noting that in Myers & Padgett 2014 and Greenwood 2016, unnaturalness stems from the articulatory difficulty of voiced obstruents while in my approach the relative naturalness of phonotactic systems with different distributions of the obstruent voicing contrast hinges on the perceptibility of the contrast in different positions. Ultimately, I will argue that only perceptual factors, not articulatory factors, bias phonological learning.

1.4.2 A P-map Implementation of Substantive Bias

The central purpose of this dissertation is to test for synchronic learning biases using artificial grammar learning; it is not to propose a model or an implementation of either complexity bias or substantive bias. In this section, though, I sketch out one possible grammatical implementation of substantive bias. To illustrate this implementation, I continue to use the example of the asymmetrical distribution of the obstruent voicing contrast across positions. In this proposal, the repository of the learner's phonetic knowledge, at least with respect to perception, is the P(erceptual)-map, or P-map (Steriade 2001b). Steriade proposed the P-map as a component of the grammar where information about perceptual similarity is stored. The P-map thus encodes the fact that voiced and voiceless obstruents are more perceptually similar in word-final position than in word-initial position (Steriade 1997; see the beginning of Chapter 3 for discussion). Then, based on perceptual information like this, the P-map projects correspondence constraints and their rankings such that constraints that militate against more perceptible changes are obligatorily ranked higher than constraints that militate against less perceptible changes.[16] More concretely, if the P-map states that the obstruent voicing contrast is more perceptible word-initially than word-finally, it will project the ranking IDENT(voice)/#__ >> IDENT(voice)/__#. That is, it is worse to change voicing at the beginning of a word than at the end of a word.

Consider now the case where the grammar contains a markedness constraint *D, which militates against voiced obstruents.[17] Given the fixed P-map-projected ranking IDENT(voice)/#__ >> IDENT(voice)/__#, three different grammars and languages can be generated.[18] If *D outranks both IDENT constraints, the language will only have voiceless obstruents. This is illustrated in the tableau in (1): only voiceless obstruents may surface in this language because voiced obstruents are banned.

---

[16] The implementation of substantive bias discussed here is couched in Standard Optimality Theory (Prince & Smolensky 1993/2004), with strict ranking of constraints. It could be straightforwardly translated to frameworks that instead use weighted constraints, such as Harmonic Grammar (Legendre, Miyata, & Smolensky 1990, Smolensky & Legendre 2006) and Maximum Entropy Harmonic Grammar (MaxEnt) (Goldwater & Johnson 2003, Hayes & Wilson 2008), by having the P-map project fixed relative weightings of constraints rather than fixed rankings. In this case, for instance, IDENT(voice)/#__ would have an obligatorily higher weight than IDENT(voice)/__# (at least in an initial learning state, if constraint weights are to be induced, as in MaxEnt). See White 2013 for a MaxEnt learning model that incorporates a P-map-based substantive bias.

[17] *D might be a member of the universal constraint set, or it might be induced by the learner through articulatory experience or language exposure. Some approaches assume that *D (or similar markedness constraints penalizing voiced obstruents) exists while *T (or similar constraints) does not, either because that is the nature of the universal constraint set or because [voice] is a privative feature and a constraint can penalize its presence but not its absence (Lombardi 1991). If markedness constraints are induced, it must be possible for constraints like *T to be learned, given the right input.

[18] For the sake of simplicity, I consider only word-initial and word-final obstruents here. Obstruents may of course occur in other positions, and the P-map accordingly contains other IDENT(voice) constraints (e.g. IDENT(voice)/V_V) whose rankings are determined by the relative perceptibility of the obstruent voicing contrast across positions.

(1)

Input	Output	*D	IDENT(voice)/#__	IDENT(voice)/__#
tab	tab	*!		
	☞ tap			*
bat	bat	*!		
	☞ pat		*	

If *D is ranked in between the IDENT constraints, the language will have a voicing contrast in obstruents word-initially but only voiceless obstruents word-finally, like final devoicing languages. The effects of this ranking are shown in the tableau in (2): obstruents may be voiced or voiceless word-initially but only surface as voiceless word-finally.

(2)

Input	Output	IDENT(voice)/#__	*D	IDENT(voice)/__#
tab	tab		*!	
	☞ tap			*
bat	☞ bat		*	
	pat	*!		

Lastly, if *D is ranked below both IDENT constraints, the language will have a voicing contrast in obstruents both word-initially and word-finally. This is illustrated in the tableau in (3): voiced and voiceless obstruents are possible in both initial and final positions.

(3)

Input	Output	IDENT(voice)/#__	IDENT(voice)/__#	*D
tab	☞ tab			*
	tap		*!	
bat	☞ bat			*
	pat	*!		

Crucially, because the P-map dictates the ranking of the two IDENT constraints relative to each other, it is impossible to generate a language that contrasts voicing in obstruents word-finally but not word-initially.[19] This gives rise to the phonotactic implicational mentioned above: if a language has an obstruent voicing contrast in word-final position, it will also have an obstruent voicing contrast in word-initial position. More generally, the P-map will project constraint rankings that ensure that if a language has a contrast in obstruent voicing in a given position, it will also have a contrast in obstruent voicing in all positions where that contrast is easier to perceive. Thus the phonetic knowledge contained in the P-map and the constraint rankings projected from it may shape what phonotactic systems learners are able, or at least prefer, to acquire.[20] The substantive bias is implemented via the P-map (see White 2013 for such an implementation of the bias against phonetically unnatural saltatory alternations).

Incorporating substantive bias into the grammar with the P-map in the way just described predicts a natural pattern of results in the AGL paradigm I use. As mentioned above, I expose participants to an obstruent voicing contrast either word-initially or word-finally and then test whether they extend the contrast to the other position. The prediction from phonetic naturalness is that participants should extend the contrast from word-final position to word-initial position more than they extend it from word-initial position to word-final position, and the P-map implementation shows why this is the case. If participants encounter the obstruent voicing contrast in word-final position, they know that *D, or whatever markedness constraint enforces neutralization of the contrast, is ranked below IDENT(voice)/__#, as in the tableau in (3).

---

[19] See Lombardi 1999 and, later, Lombardi 2001 for a mechanically similar analysis of laryngeal neutralization that also captures this typology (and in which [voice] is privative).

[20] The analysis sketched out here in Standard Optimality Theory predicts a “hard bias” whereby phonotactic systems that do not conform to the phonotactic implicational cannot be learned at all (and phonotactic systems that do conform are learned perfectly). Like other AGL experiments, my experiments test for a “soft bias” (cf. White 2013): unnatural systems may still be learned to a degree but are predicted to be learned worse than natural systems. Additionally, AGL participants’ behavior exhibits gradience, not perfect learning. Modeling gradient, rather than categorical, results requires a framework of the phonological grammar that can generate variation, such as MaxEnt (Wilson 2006, White 2013; see Footnote 16).

Additionally, by a P-map-projected fixed ranking, IDENT(voice)/__# is ranked below IDENT(voice)/#__, so by transitivity *D is also ranked below IDENT(voice)/#__ (also illustrated in (3)). This ranking, IDENT(voice)/#__ >> *D, ensures that the voicing contrast will not be neutralized in word-initial position, so even if participants have not been exposed to the voicing contrast word-initially they should assume that it exists there and accept voiced and voiceless word-initial obstruents in the test phase.

On the other hand, if participants encounter the obstruent voicing contrast in word-initial position, all they know is that IDENT(voice)/#__ outranks *D (as in the tableau in (2) or the tableau in (3)). From the P-map, they already have the ranking IDENT(voice)/#__ >> IDENT(voice)/__#, but they do not have information that would allow them to determine the relative ranking of *D and IDENT(voice)/__#. According to a view commonly adopted in the phonological acquisition literature, markedness constraints outrank faithfulness constraints in the initial learning state and are only demoted if the learner encounters evidence that they can be violated (Smolensky 1996 and references therein). Given that participants have no evidence that *D is ranked below the constraint that preserves the voicing contrast in word-final position, they should think that the obstruent voicing contrast does not exist there and accept only voiceless word-final obstruents in the test phase (i.e. they should adopt the grammar in (2)).

This section is merely intended to give an idea of how the substantive bias effects tested in the phonotactic learning experiments in this dissertation might be modeled in the grammar if they are shown to be active. I do not claim that this is the only way substantive bias could constrain phonological learning.

1.4.3 Outline of the Dissertation

The remainder of the dissertation is organized by experiment. Chapter 2 reports on Experiments 1 and 2, which test whether learners reproduce the phonotactic implicational whereby place of articulation contrasts in word-final stops entail place of articulation contrasts in word-initial stops but place of articulation contrasts in word-initial stops do not necessarily entail contrasts in word-final stops. In Experiment 1, I exposed participants to languages in which labial, coronal, and velar stops occurred in word-initial or word-final position while only coronal stops occurred in the other word-edge position. I then tested whether participants extended the place of articulation contrasts they had heard in one position to the other position. The prediction from substantive bias was that participants would extend the contrasts more from word-final to word-initial position than from word-initial to word-final position, but no such asymmetry emerged. Instead, participants extended the contrasts to the same degree in both directions, showing no effect of substantive bias. Experiment 2 was a modified version of Experiment 1 designed to encourage more implicit learning, and it too uncovered no evidence for substantive bias.

Chapter 3 reports on Experiments 3 and 4, which test a different phonotactic implicational: an obstruent voicing contrast in word-final position entails an obstruent voicing contrast in word-initial position, but the reverse is not necessarily true. In Experiment 3, I exposed participants to a stop voicing contrast in one word-edge position while in the other word-edge position stops were either only voiced or only voiceless. Substantive bias predicts that participants should extend the stop voicing contrast from word-final to word-initial position more than from word-initial to word-final position. The results did not fulfill this prediction, yielding no support for substantive bias, but extension of the stop voicing contrast to the opposite word-edge position did vary between conditions. I argue that the observed differences are the result of a complexity bias that was active because the sonorant filler consonants in the experimental stimuli rendered certain conditions more complex than others. Experiment 4 again tested whether participants would reproduce the phonotactic implicational, but the filler consonants were changed to voiceless fricatives to reverse the predictions of complexity bias. In Experient 4, participants did extend the stop voicing contrast more from word-final to word-initial position than in the other direction, a pattern that constitutes evidence for substantive bias. Additionally, as in Experiment 3, there was an effect of complexity bias. Together, the complexity bias effects found in Experiments 3 and 4 demonstrate how the broader phonological structure of an artificial language, including the properties of non-critical, filler sounds, affects performance. Participants seem able to infer phonotactic constraints relative to the experiment-internal distribution of sounds, opting for the simplest constraint with which they can master the pattern.

Chapter 4 reports on Experiment 5, which unlike Experiments 1-4 is an alternation learning experiment. Experiment 5 tested for substantive bias by comparing the learnability of final devoicing, a natural and typologically frequent phonological process, and final voicing, the complementary process, which is unnatural and unattested. Contrary to expectations, unnatural final voicing was learned better than natural final voicing; the experiment thus did not provide evidence for substantive bias. I argue for a complexity bias account of the results, analogous to the account provided for the phonotactic learning results of Experiment 3. Experiment 5 also tested the learning of a more complex exchange pattern combining final voicing and devoicing. Participants learned this alternation surprisingly well, despite its complexity, but I attribute this to exposure to a greater proportion of alternating forms.

Chapter 5 concludes the dissertation by synthesizing the five experiments and offering a new interpretation of their results and the wider substantive bias literature. I suggest that substantive bias must be broken down into articulatorily-grounded and perceptually-grounded naturalness biases and that only perceptually-grounded bias may constrain phonological learning. Finally, I outline some future directions.

CHAPTER 2: POSITIONAL EXTENSION OF MAJOR PLACE CONTRASTS

The first phonotactic implicational I tested concerns the positional distribution of major place contrasts in consonants. If a language has major place contrasts in stops post-vocalically/in coda position (e.g. /ap/ vs. /at/ vs. /ak/), it should also have major place contrasts in stops pre-vocalically/in onset position (e.g. /pa/ vs. /ta/ vs. /ka/), but not necessarily vice versa (Blevins 2004, O'Hara 2019). This implicational is phonetically motivated. Based on Fujimura, Macchi & Streeter's (1978) finding that listeners attend more to CV transitions than to VC transitions in identifying the place of articulation of voiced stops, Steriade (1994, 2001a) proposed that differences in place of articulation are more perceptible pre-vocalically (_V) than post-vocalically (V_).[21] Blevins (2004) also synthesizes research demonstrating that the strongest cues to place of articulation in stops are in the CV transition and that place cues are weaker when the stop is not followed by a vowel. In other words, consonants differing in place of articulation are more perceptually similar (i.e. harder to distinguish) after a vowel than before a vowel. If a place contrast exists post-vocalically, where it is harder to perceive, then it should also exist pre-vocalically, where it is easier to perceive. It would be phonetically unnatural for a language to exhibit a contrast in a less salient position, perceptually speaking, while not exhibiting it in a more salient one.

The phonotactic implicational about the distribution of major place contrasts appears to hold cross-linguistically. Steriade (1994) observes that place neutralization typically occurs in coda position; similarly, Lombardi (2001) notes that many languages exhibit neutralization to an

---

[21] More precisely, this applies to major place contrasts (labial vs. coronal vs. dorsal) in oral stops. The following discussion refers only to these contrasts. Other contrasts, such as that between retroflex and non-retroflex consonants, have different cues and therefore different patterns of perceptual similarity across positions (see Blevins 2004, Steriade 2001a).

unmarked place in coda position. Further typological evidence for the preference for maintaining major place contrasts pre-vocalically comes from the fact that coda consonants are more likely to be targets of place assimilation than onset consonants (Jun 1995). Referring specifically to word edges, Blevins (2004:117) concludes, based on observations from dozens of genetically diverse languages, that "word-final neutralization of major place features is common, while word-initial neutralization of the same features is unattested." This statement is perfectly consistent with the phonotactic implicational. O'Hara (in prep.) developed a database examining the word-initial and word-final consonant inventories in about 170 language from over fifty language families. Of the 57 languages in the database that exhibit exactly [p t k] word-finally, all also allow [p t k] (and sometimes other stops besides) word-initially (O'Hara 2019). On the other hand, of the 110 surveyed languages that exhibit exactly [p t k][22] word-initially, about half allow fewer stops in word-final position (with by far the most common pattern being a ban on all three stops word-finally) (O'Hara in prep.). This typological data also supports the implicational whereby major place contrasts in word-final position entail those same contrasts word-initially but not vice versa.

2.1 Experiment 1: Positional Extension of Major Place Contrasts

2.1.1 Method

If phonetic knowledge influences learners' acquisition of phonological patterns and learners are biased toward phonetically natural phonological systems, they should make

---

[22] In conducting his typological study, O'Hara looked at the stop series with the greatest number of licit places of articulation in a given language, which was most often the plain voiceless stop series. Thus when he classifies a language as having exactly [p t k] word-initially, he means it has maximally a three-way major place distinction word-initially; that the language may also have e.g. the voiced stops [b d g] is not excluded.

inferences about language in accordance with phonotactic implicationals like the one above. To test whether this is the case for the major place contrast implicational, I exposed participants to artificial languages featuring a three-way place contrast in either word-initial (pre-vocalic) or word-final (post-vocalic) position and tested whether they assumed the contrast existed in the other position as well. If participants behave in a way that is consistent with the phonotactic implicational, they will extend the place contrast more from word-final to word-initial position than from word-initial to word-final position.

2.1.1.1 Conditions

There were two training conditions, Initial and Final. In the Initial condition, participants heard words beginning with labial, coronal, and velar stops (e.g. ***p****ilan*, ***t****ulir*, ***k****ilun*), but in word-final position stops were only ever coronal (e.g. *wiru****t***). In the Final condition, participants heard words ending with labial, coronal, and velar stops (e.g. *ruja****p***, *wiru****t***, *wanu****k***), but word-initial stops were always coronal (e.g. ***t****ulir*). I chose coronal place of articulation for the position in which stops do not contrast in place because coronal place is commonly considered unmarked (Kean 1975, Paradis & Prunet 1991, de Lacy 2006, Rice 2007). That said, Blevins (2004) disputes the special status of coronal place as the unmarked place and notes that there is little support for this proposal from the typology of place neutralization, given that neutralization of major place contrasts to coronal place is rarely attested. Blevins seems to be referring principally to diachronic sound changes, but O’Hara’s (in prep.) typological study of the distribution of major place contrasts also supports this claim. Of the 110 languages he surveys that have just a three-way major place contrast in stops word-initially, as in the Initial condition of my experiment, only one, Finnish, allows just [t] word-finally. The vast majority of the languages either have all three places of articulation in stops word-finally as well or have no stops word-finally. Thus the pattern exhibited in the artificial language of the Initial condition is in fact typologically quite rare, even though the word-final place neutralization is common. Diachronically, at least, when labial, coronal, and dorsal oral stops all neutralize it is generally not to one of those three places but rather to a glottal stop (Blevins 2004). Lombardi (2001) notes that synchronic neutralization of place distinctions to glottal place (/h/ or /ʔ/) is also common, and she proposes that pharyngeals, which for her include /h/ and /ʔ/, are even less marked than relatively unmarked coronals. Since my experiment participants were native English speakers, however, I used only phonemes of English in the stimuli, which meant that the three-way major place contrast had to neutralize to labial, coronal, or velar place. Of these options, I chose coronal place in light of the literature on the unmarked status of coronals.

The Initial language is natural in that it features major place contrasts in stops word-initially but not word-finally, a pattern consistent with the phonotactic implicational. The Final language, on the other hand, is unnatural: in having major place contrasts word-finally but not word-initially, it violates the implicational. Table 2.1 shows which types of stops occurred in which positions in the two training conditions.

Table 2.1: Experiment 1 Training Conditions

	#P	**#T**	**#K**	**P#**	**T#**	**K#**
Initial	✓	✓	✓	✗	✓	✗
Final	✗	✓	✗	✓	✓	✓

The test items were identical in the two conditions. In the test phase, both Initial and Final participants heard words beginning and ending with labial, coronal, and velar stops.

2.1.1.2 Materials

The training and test items were all nonce words of the shape $C_1VC_2VC_3$. Either C_1 or C_3 was a stop drawn from [p t k b d ɡ]. The other two Cs were sonorants drawn from [n l ɹ j w]. I excluded [m] from these “filler” consonants so that the position without place contrasts in stops would not feature a place contrast in another type of non-continuant consonant. The glides [j] and [w] did not occur word-finally, and no item contained a [ji] or [wu] sequence. In each word, all three Cs were different. The vowels were drawn from [i ɑ u]. Labial, coronal, and velar stops were equally represented in the position with the place contrast in training and across positions in test. Additionally, voiced and voiceless stops were equally represented across positions and places of articulation in both the training and test phases. In the training phase, of the items with stops in the position with the place contrast (e.g. items beginning with stops in the Initial condition), half were members of minimal triplets for place of articulation (e.g. ***b**awan*-***d**awan*-***g**awan* in the Initial condition). This was to encourage participants to notice that place of articulation was contrastive in stops in the given position. In the test phase, half of all items were members of minimal triplets for place of articulation. In both training and test, half of the items were iambs and half were trochees. Stress did not correlate with the position that featured the place contrast or with the syllable in which the stop (always either word-initial or word-final) occurred. This was to prevent participants from associating the place contrast with stress instead of position and to prevent stress from drawing undue attention to the stops. Stress was also not correlated with the voicing of the stop. Table 2.2 shows some sample training items from the Initial condition. The full set of training and test stimuli for Experiment 1 are given in the Appendix.

Table 2.2: Sample Training Items in the Initial Condition

#P	#T	#K	P#	T#	K#
pínir **b**ilún …	**t**ínir **d**irúl …	**k**ínir **g**urúl …		jáwi**t** rujú**d** …	

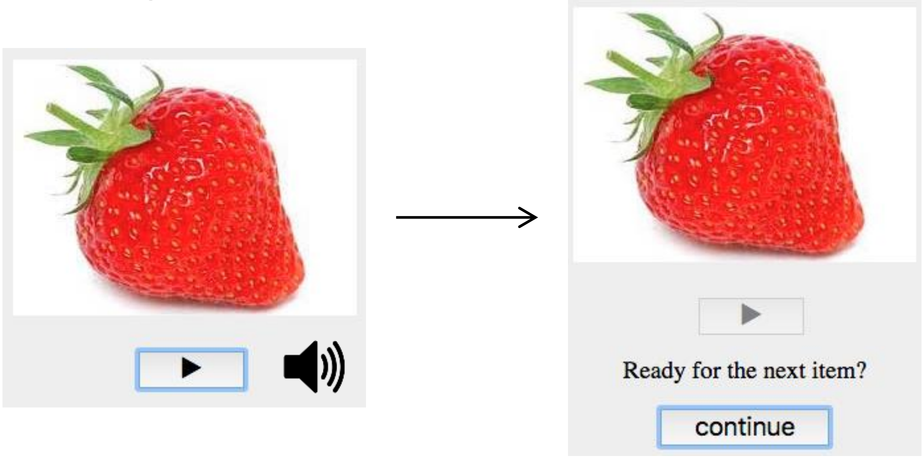
The stimuli were recorded by a phonetically-trained male native speaker of American English who was naïve to the purpose of the experiment. Voiced stops were fully voiced, voiceless stops were aspirated, and word-final stops were released. Unstressed vowels were not reduced. Otherwise, pronunciation was as in American English. The stimuli were produced in isolation as if each item was a sentence unto itself. The stimuli were recorded in a sound-attenuated room using a head-mounted microphone. Recording was done using Audacity with a sampling rate of 22,050 Hz.

2.1.1.3 Procedure

The experiment was conducted online using Experigen (Becker & Levine 2013). Participants were instructed to wear headphones and to do the experiment in a quiet room. They saw written instructions informing them that they would be listening to some words of a new language. They would then be presented with additional words and asked whether those words sounded like they could also be from the language they heard in the first part of the experiment. Before the experiment proper began, they were prompted to play two sound files and type in the English word they had heard. The first word, *pad*, could be played multiple times so that participants could adjust the volume on their computer to a satisfactory level. The second word, *bat*, could only be played once. These test words were included to try to ensure participants were doing the experiment under acceptable listening conditions.

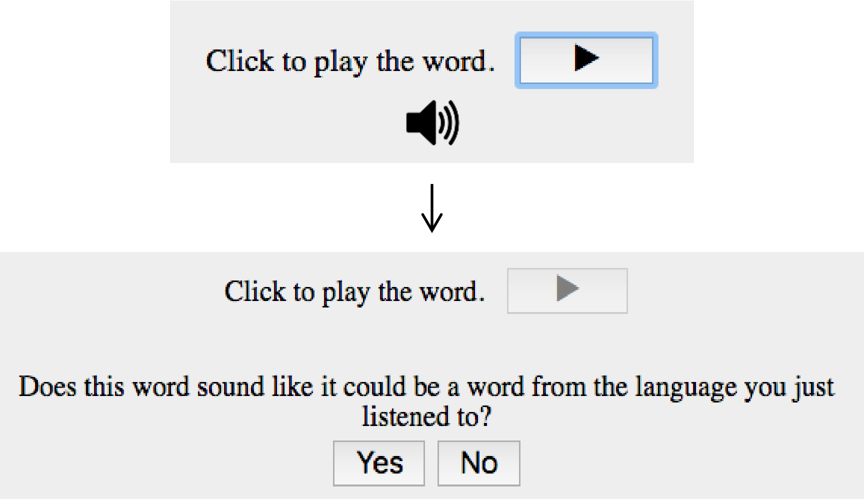
The experiment began with the training phase (see Figure 2.1). Participants proceeded at their own pace through two blocks of the same 40 training items. The order of the training items was randomized within each block for each participant, and there was no interruption between the two blocks. In each training trial, participants saw an image and clicked a button to hear the word for that image. Each sound file could only be played once. Participants then clicked to continue to the next training trial. They were encouraged to say the words out loud to help them learn the language. After the training phase, written instructions appeared telling participants that they would hear some additional words and should make their best guess as to whether each word sounded like it could also be a word from the language they had just listened to. There were no images in the test phase (see Figure 2.2). In each test trial, participants clicked a button to hear the sound file. They then had to click Yes or No to indicate whether they thought the word sounded like it could be from the language they had just listened to. There was a single test block consisting of 48 test items, which were the same in both conditions. Their order was randomized for each participant. At the end of the experiment, participants answered questions about their language background, whether they had noticed a pattern in their training language, and whether they had used a particular strategy in the test phase.

Figure 2.1: Event Sequence for One Training Trial



Ready for the next item?

Figure 2.2: Event Sequence for One Test Trial



Click to play the word.

Click to play the word.

Does this word sound like it could be a word from the language you just listened to?

There were three types of items in the test phase. While the same set of test items was used in both conditions, which types test items fell into depended on the condition. *Familiar Conforming* items contained stops whose place of articulation and position conformed to the trained pattern (e.g. #P, #T, #K, and T# in the Initial condition) and were words that were heard in training. *Novel Conforming* items also contained stops whose place of articulation and position conformed to the trained pattern, but these words had not been heard in training. Finally, *Novel Nonconforming* items featured the place of articulation and position combinations not heard in training (e.g. P# and K# in the Initial condition). Table 2.3 gives sample test items of each type for the two training conditions.

Table 2.3: Sample Test Items for Each Training Condition in Experiment 1

	Familiar Conforming	**Novel Conforming**	**Novel Nonconforming**
Initial	**p**ínir	**p**anúl	nálu**p**
Final	rujá**p**	nálu**p**	**p**anúl

2.1.1.4 Participants

The participants were native English speakers recruited through the UCLA Psychology Subject Pool. I excluded participants who reported that they were not in fact native English speakers (12 participants), had taken more than one linguistics class (7 participants), reported a history of speech or hearing impairments (4 participants), gave an incorrect response to either of the two test words that preceded the experiment (0 participants in this experiment), or accepted all test items (2 participants). After exclusions (25 out of 74 participants), there were 25 participants in each condition.

2.1.2 Predictions

While not central to the research question, a first prediction is that the acceptance rate of Familiar Conforming items should be greater than the acceptance rate of Novel Conforming items. In other words, participants should accept words they heard in the training phase more often than they accept words that fit their training pattern but are new to them (cf. Linzen & Gallagher 2017).

Participants' acceptance rates of Novel Nonconforming items, relative to Novel Conforming items, indicate whether they have extended the place contrast in stops to a new position in a given condition. For instance, if participants in the Initial condition accept test items with word-final labial and velar stops (P# and K#), they have extended the word-initial place contrast they encountered in training to word-final position. Recall the phonotactic implicational: if a language has major place contrasts word-finally, it should also have major place contrasts word-initially, but not necessarily the other way around. Performance consistent with the implicational would be asymmetric extension: participants exposed to the place contrast word-finally should extend it to word-initial position more than participants exposed to the place contrast word-initially extend it to word-final position. Extension of the place contrast to the other position manifests as erroneously accepting Novel Nonconforming items, so participants trained on the contrast word-finally should more readily accept their Novel Nonconforming items (#P and #K) than participants trained on the contrast word-initially accept their Novel Nonconforming items (P# and K#). Assuming similar performance in both conditions on Novel Conforming items, meaning that participants in both conditions learned to an equal degree what types of words did belong to their language, a higher acceptance rate of Novel Nonconforming items in the Final condition relative to the Initial condition would constitute evidence that learners are biased toward phonetically natural phonotactic systems. Put another way, if learning is substantively biased, participants should learn the unnatural Final language worse than the natural Initial language, as demonstrated by their mistakenly “filling in” the missing initial labial and velar stops in their language. If learners are not biased toward phonetically natural phonotactic systems, participants should accept Novel Nonconforming items at similar rates in both conditions since the two training patterns are of equal formal complexity.

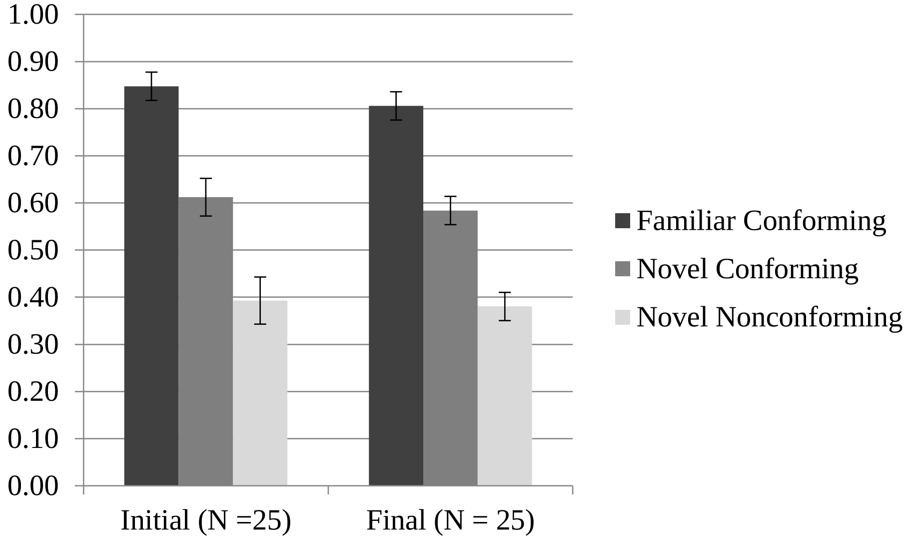
2.1.3 Results

Figure 2.3 shows the acceptance rates of the three types of test items in each condition.[23] At a glance, it appears that participants accepted Familiar Conforming items more than Novel Conforming items and Novel Conforming items more than Novel Nonconforming items. The rates at which participants accepted each type of test item do not seem to differ between the two conditions. In other words, it looks as though participants behaved identically in both conditions.

---

[23] Here and throughout error bars represent ±1 standard error.

Figure 2.3: Acceptance of Test Items by Condition



I analyzed the results with mixed-effects logistic regressions using the `glmer()` function from the *lme4* package (Bates et al. 2015) in R. Because the factors of Familiarity and Conformity are not fully crossed, I carried out separate analyses of the conforming test items (both familiar and novel) and the novel test items (both conforming and nonconforming) (cf. Linzen & Gallagher 2017). I first fit a mixed-effects logistic regression to the conforming items (Familiar Conforming and Novel Conforming) with response (accept or reject) as the dependent variable, Familiarity (familiar vs. novel), Condition (Initial vs. Final), and their interaction as fixed effects, random intercepts for subject and item, and by-subject random slopes for Familiarity.[24] Table 2.4 gives the fixed effects for this model. There was a significant main effect of Familiarity: Familiar Conforming items were accepted more often than Novel Conforming items ($\beta = 1.526$; $p < 0.001$). This is as predicted: participants should be better at accepting conforming items they heard in the training phase than conforming items that are new. There was

---

[24] I did not include by-item random slopes for Condition because although the test items were the same in both conditions, they did not fall into the same types in each condition. Thus the Familiar Conforming items in the Initial and Final conditions are not exactly the same sets of items, and the same is true for the Novel Conforming items.

no significant effect of Condition, nor was there a significant interaction of Familiarity and Condition.

Table 2.4: Experiment 1—Fixed Effects of the Conforming Items Model

	β	p
Intercept	0.547	0.011*
Familiarity = familiar (vs. unfamiliar)	1.526	<0.001***
Condition = Final (vs. Initial)	-0.167	0.529
Familiarity × Condition	-0.157	0.643

I then fit a mixed-effects logistic regression to the novel items (Novel Conforming and Novel Nonconforming) with response (accept or reject) as the dependent variable, Conformity (conforming vs. nonconforming), Condition (Initial vs. Final), and their interaction as fixed effects, random intercepts for subject and item, and by-subject random slopes for Conformity. Table 2.5 gives the fixed effects for this model. There was a significant main effect of Conformity: Novel Conforming items were accepted more often than Novel Nonconforming items ($\beta = 1.094$; $p < 0.001$). Of words they had never heard before, participants in both conditions preferred conforming items to nonconforming items. There was no significant effect of Condition, and, crucially, there was no significant interaction of Conformity and Condition. It is in this interaction that an effect of substantive bias would emerge: if learners favor phonetically natural phonotactic systems, there should be an interaction of Conformity and Condition such that participants in the Final condition accepted Novel Nonconforming items at a higher rate than participants in the Initial condition. Since this interaction is not significant, the results do not support substantive bias. Instead, participants learned the natural Initial language and unnatural Final language equally well, exhibiting no difference in the degree to which they "filled in" labial and velar stops in the position in which they had been withheld in training.

Table 2.5: Experiment 1—Fixed Effects of the Novel Items Model

	Coefficient	***p***
Intercept	-0.520	0.019*
Conformity = conforming (vs. nonconforming)	1.094	<0.001***
Condition = Final (vs. Initial)	-0.057	0.853
Conformity × Condition	-0.116	0.744

The acceptance rates of Novel Conforming items were significantly above chance in both conditions (though only barely in the Final condition), meaning that both Initial and Final participants correctly generalized to new words in their language. The acceptance rates of Novel Nonconforming items were significantly below chance in both conditions, meaning that both Initial and Final participants correctly rejected words not in their language.[25] In other words, participants in both conditions did not fill in the place of articulation gaps in their training languages; they instead learned the gaps. The lack of interactions in the regression models means that there was no difference in participants' ability to generalize to new words or reject nonconforming words between the conditions.

2.1.4 Discussion

The participants in Experiment 1 did not reproduce the phonotactic implicational about the positional distribution of major place contrasts. That is, they did not extend the place contrast more from word-final to word-initial position than from word-initial to word-final position. Thus the experimented yielded no evidence for substantive bias.

---

[25] Whether acceptance rates were significantly above chance was determined by fitting mixed-effects logistic regressions to the novel items (Novel Conforming and Novel Nonconforming) with the same random effects structure as in the model in Table 2.5 and changing the reference levels of the factors Condition and Conformity so that in each of the four models the intercept represented the acceptance rate for a different combination of Condition and Conformity. When the intercept represented Initial Novel Conforming or Final Novel Conforming items, it was significantly above chance (Initial: $\beta = 0.575$, $p = 0.004$; Final: $\beta = 0.402$, $p = 0.042$). When the intercept represented Initial Novel Nonconforming or Final Novel Nonconforming items, it was significantly below chance (Initial: $\beta = -0.520$, $p = 0.019$; Final: $\beta = -0.577$, $p = 0.009$).

One reason no substantive bias effect emerged could be that participants' learning was too explicit because the pattern was too easy to identify. Moreton & Pertsova (2016) demonstrate that learning in artificial phonotactic learning experiments can be more or less explicit depending on an experiment's design and procedure. Training with feedback, instructing participants to try to find a rule, and using "easily verbalizable" features all foster explicit learning while training without feedback, not mentioning rules in the instructions, and using features that are not easy to verbalize foster implicit learning (Moreton & Pertsova 2016:277). In Experiment 1, there was no feedback in the training phase, as the training phase did not ask participants to make any kind of decision. Nor was there any feedback in the test phase, when participants did have to choose between a correct and an incorrect response on each trial. Neither the training instructions nor the test instructions made any mention of rules. The feature of position (initial or final) is conceivably fairly easy to verbalize even for non-linguists, but segmental phonological features, including place of articulation, are likely much harder to verbalize. From the perspective of design, then, Experiment 1 looks more like experiments that attempt to foster implicit learning.

Moreton & Pertsova (2016) conducted an experiment in which participants had to learn to distinguish masculine and feminine words in an artificial language in which some phonological or semantic property was a perfect cue to gender (e.g. disyllabic words are masculine and trisyllabic words are feminine). There were two conditions: Feedback, whose design and procedure were meant to foster explicit learning, and No-Feedback, whose design and procedure were meant to foster implicit learning. Signs that participants are engaging in explicit learning include reporting seeking a rule and identifying the correct rule (Moreton & Pertsova 2016:278). In the Feedback condition, 82% of participants reported seeking a rule during the training phase, and 61% reported using a rule to make their decisions during the test phase. In the

No-Feedback condition, 56% of participants reported seeking a rule in training, and 43% reported using a rule in test. In my Experiment 1, participants were only asked about their strategy once, in the survey that followed the test phase. Most participants (17 out of 25 (68%) in the Initial condition and 19 out of 25 (76%) in the Final condition) reported seeking a rule (defined as stating a connection between their responses and some property of the stimuli) to distinguish words that were in their language from words that were not in their language. The remaining participants reported a more intuitive strategy.[26] It is difficult to directly compare these survey results to those in Moreton & Pertsova's experiment due to differences in the training procedure and in the wording of the survey questions, but the extent of rule-seeking in Experiment 1 looks as though it lies in between the extent of rule-seeking in their Feedback (explicit learning) and No-Feedback (implicit learning) conditions. In their experiment, 14% of No-Feedback participants identified the correct rule while 29% of Feedback participants identified the correct rule. In Experiment 1, two participants (8%) found the correct rule in the Initial condition and four participants (16%) found the correct rule in the Final condition, where the correct rule was defined as one that mentioned labials and/or velars *and* position. In terms of the proportion of participants that identified the correct rule, Experiment 1 looks more like Moreton & Pertsova's No-Feedback condition, which was designed to elicit implicit learning.

---

[26] I tested for the substantive bias effect among the non-rule-seeking participants only (14 total; 8 in the Initial condition and 6 in the Final condition). The raw results appear promising: non-rule-seeking participants' mean acceptance rates of Novel Conforming items were similar (59% in the Initial condition and 61% in the Final condition), but the mean acceptance rate of Novel Nonconforming items was higher in the Final condition (47%) than in the Initial condition (37%). This looks like greater extension of the place contrasts from final position to initial position than vice versa. However, when I fit a mixed-effects logistic regression to the novel items (the same type of model as in Table 2.5), the interaction of Conformity × Condition that would demonstrate asymmetric extension and therefore support substantive bias was not significant ($\beta = 0.566$; $p = 0.323$). This is likely due to the small number of participants falling in the implicit learning category.

While Experiment 1 already incorporates design properties previously used to foster implicit learning and partly resembles Moreton & Pertsova's No-Feedback (implicit learning) condition in its results, Moreton & Pertsova demonstrated that participants may engage in explicit learning even in an experiment meant to encourage implicit learning. They also showed that changing properties of an experiment can affect what kind of learning participants engage in. Thus, even if Experiment 1 elicited a certain amount of implicit learning, it should still be possible to make learning even more implicit. I hypothesized that making the task harder might foster more implicit learning, which might in turn cause substantive bias to emerge. To test this, I carried out a modified version of the positional extension of place contrasts experiment.

2.2 Experiment 2: Positional Extension of Major Place Contrasts with Fillers

2.2.1 Method

Experiment 2 was identical to Experiment 1 except for three differences. First, I added 20 filler items in which all three consonants were sonorants (e.g. *lanir*) to the training phase. With the 40 critical items from Experiment 1, this yielded a total of 60 training items, one third of which were fillers. The all-sonorant fillers were intended to disguise the pattern, making it less obvious and thus leading to more implicit learning. The filler items were recorded by the same male speaker who recorded the stimuli for Experiment 1. The full set of training and test stimuli for Experiment 2 are given in the Appendix. Second, I increased the number of training blocks from two to three. This was done because the acceptance rate of Novel Conforming items in the Final condition was just barely above chance in Experiment 1. Acceptance rates of Novel Conforming items must be above chance to show that participants have learned to accept the types of items they were trained on. Without this demonstration of learning, performance on

Novel Nonconforming items is harder to interpret. I therefore increased the number of training blocks in Experiment 2 to try to push Novel Conforming acceptance rates up a little.

Finally, I removed the Familiar Conforming items from the test phase. This was done because after implementing the two changes above, pilot data showed no increase in the acceptance rates of Novel Conforming items. In particular, increasing the number of training blocks from two to three was not having the desired effect of boosting Novel Conforming acceptance rates. The presence of the Familiar Conforming items in the test phase may have been keeping down the acceptance rates of Novel Conforming items, since there was a class of test items that were even better (i.e. more acceptable) than the Novel Conforming items, namely, the Familiar Conforming items. In other words, the inclusion of test items repeated from training may have imposed a ceiling on the acceptance rate of items that still fit the training pattern but had not been encountered before. As the results will show, eliminating the Familiar Conforming items did lead to higher acceptance rates for Novel Conforming items.

In all other respects, Experiment 2 was an exact replication of Experiment 1.

2.2.1.1 Participants

The participants were native English speakers recruited through the UCLA Psychology Subject Pool. I applied the same exclusion criteria as in Experiment 1 and excluded 25 out of 67 participants. Of the 25 excluded participants, 10 were not in fact native English speakers, 7 had taken more than one linguistics class, 1 reported a history of speech or hearing impairments, and 7 gave an incorrect response to either of the two test words that preceded the experiment. After exclusions, there were 21 participants in each condition. None of the participants in Experiment 2 had participated in Experiment 1.

2.2.2 Predictions

The predictions for Experiment 2 are the same as the predictions for Experiment 1, with the exception that Familiar Conforming items are not predicted to be accepted more than Novel Conforming items because there are no Familiar Conforming items. If learners are biased toward phonetically natural phonotactic systems, participants in the Final condition should more readily accept Novel Nonconforming items than participants in the Initial condition. This would show that learners are more likely to assume place contrasts exists word-initially if they have been taught they exist word-finally than they are to assume place contrasts exist word-finally if they have been taught they exist word-initially. If there is no substantive bias at work, acceptance rates of Novel Nonconforming items should not differ between the two conditions.

2.2.3 Results

Figure 2.4: Acceptance of Test Items by Condition

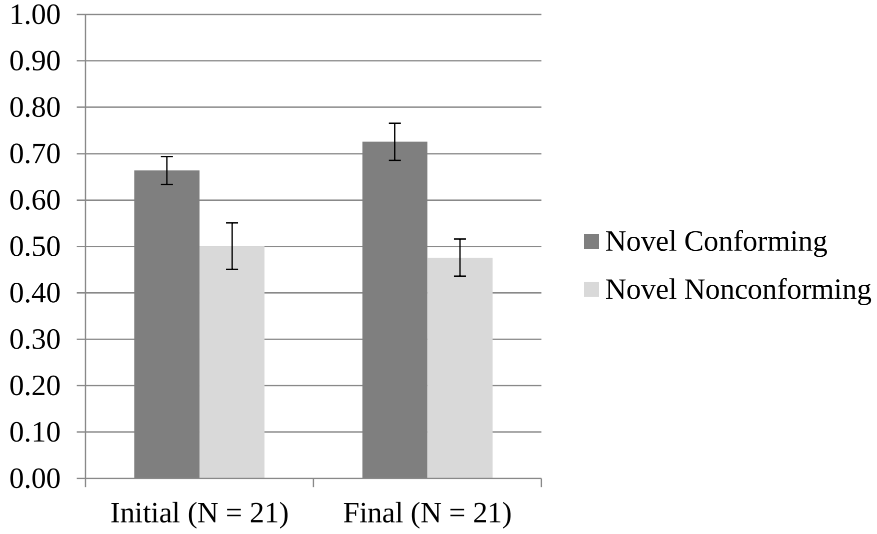


Figure 2.4 shows the acceptance rates of the two types of test items, Novel Conforming and Novel Nonconforming, in each condition. It appears that conforming items were accepted at

a higher rate than nonconforming items but that the two conditions did not otherwise differ greatly. In particular, participants in the Final condition did not accept Novel Nonconforming items more than participants in the Initial condition did.

As mentioned above, removing the Familiar Conforming items from the test phase caused acceptance rates of both Novel Conforming and Novel Nonconforming items to increase relative to Experiment 1, which did include Familiar Conforming items. In Experiment 1, the mean acceptance rate of Novel Conforming items was 61% in the Initial condition and 58% in the Final condition while in Experiment 2 the mean acceptance rate of Novel Conforming items was 66% in the Initial condition and 73% in the Final condition. Thus eliminating the Familiar Conforming test items had the desired effect of pushing up acceptance rates of Novel Conforming items. In Experiment 2, these acceptance rates were significantly above chance in both conditions.[27] Participants were able to generalize to new words that matched the pattern they had been trained on. For Novel Nonconforming items, the mean acceptance rates in Experiment 1 were 39% in the Initial condition and 38% in the Final condition while the mean acceptance rates in Experiment 2 were 50% in the Initial condition and 48% in the Final condition. Eliminating the Familiar Conforming test items also had the effect of pushing up acceptance rates of Novel Nonconforming items. In Experiment 1, Novel Nonconforming were accepted at rates significantly below chance in both conditions, but in Experiment 2, the acceptance rates of Novel Nonconforming items did not differ from chance in either condition.[28]

---

[27] This was determined by fitting a mixed-effects logistic regression to the test items (Novel Conforming and Novel Nonconforming) and changing the reference levels of the factors Condition and Conformity so that the model intercept represented the acceptance rate for the relevant combination of Condition and Conformity. When the intercept represented Initial Novel Conforming items or Final Novel Conforming items, it was significantly above chance (Initial: $\beta = 0.739$, $p < 0.001$; Final: $\beta = 1.028$, $p < 0.001$).

[28] This was determined by fitting a mixed-effects logistic regression to the test items (Novel Conforming and Novel Nonconforming) and changing the reference levels of the factors Condition and Conformity so that the model

In other words, while participants in Experiment 1 were able to correctly reject words not in their language at above chance levels, participants in Experiment 2 were not able to do so. These participants did not go as far as filling in the place of articulation gaps in their training languages; rather, they did not know whether those gaps were there or not and behaved as though they were guessing in the test phase.

To test for asymmetric positional extension of the place contrasts, I fit a mixed-effects logistic regression to the test items (Novel Conforming and Novel Nonconforming) with response (accept or reject) as the dependent variable, Conformity (conforming vs. nonconforming), Condition (Initial vs. Final), and their interaction as fixed effects, random intercepts for subject and item, and by-subject random slopes for Conformity. Table 2.6 gives the fixed effects for this model. There was a significant main effect of Conformity: Novel Conforming items were accepted more often than Novel Nonconforming items ($\beta = 0.771$; $p = 0.022$). As in Experiment 1, participants preferred novel words that fit the pattern they had been trained on to novel words that did not. There was no main effect of Condition and no interaction of Conformity and Condition.

Table 2.6: Experiment 2—Fixed Effects of the Model

	Coefficient	***p***
Intercept	-0.032	0.904
Conformity = conforming (vs. nonconforming)	0.771	0.022*
Condition = Final (vs. Initial)	-0.031	0.935
Conformity × Condition	0.319	0.524

---

intercept represented the acceptance rate for the relevant combination of Condition and Conformity. When the intercept represented Initial Novel Nonconforming items or Final Novel Nonconforming items, it did not differ significantly from chance (Initial: $\beta = -0.032$, $p = 0.904$; Final: $\beta = -0.063$, $p = 0.814$).

2.2.4 Discussion

The lack of an interaction of Conformity and Condition means that, like in Experiment 1, there was no asymmetric extension of the place contrasts in Experiment 2. Participants in the Final condition did not extend the place contrasts more from word-final to word-initial position than participants in the Initial condition extended the place contrasts from word-initial to word-final position. Experiment 2 thus did not yield evidence for substantive bias either. Though the actual acceptance rates of Novel Conforming and Novel Nonconforming items changed from Experiment 1 to Experiment 2 due to the removal of the Familiar Conforming items, the natural Initial language and the unnatural Final language were still learned equally well.

The main motivation for running Experiment 2 was to try to make participants' learning more implicit and thereby allow a substantive bias to emerge. To foster implicit learning, I added filler items to the training phase in hopes of making the distribution of labial, coronal, and velar stops less obvious. One might ask whether this modification actually caused more implicit learning in Experiment 2 as compared to Experiment 1. In Experiment 2, roughly half of participants (10 out of 21 (48%) in the Initial condition and 11 out of 21 (53%) in the Final condition) reported seeking a rule to distinguish words in their language from those not in their language. Only one participant in each condition found the correct rule. In Experiment 1, two thirds (Initial condition) or three quarters (Final condition) of participants reported seeking a rule, so there does seem to have been somewhat more implicit learning in Experiment 2. This did not lead to the emergence of substantive bias, however. The natural and unnatural patterns were still learned to an equal degree.

Together, Experiments 1 and 2 show that the phonotactic implicational whereby major place contrasts in post-vocalic position entail major place contrasts in pre-vocalic position, but not vice versa, is not reproduced in an artificial grammar learning paradigm. Instead, participants learned equally well a natural language in which place was contrasted in stops only word-initially and an unnatural language in which place was contrasted in stops only word-finally. Neither experiment's results provide support for substantive bias. In Experiment 2, I tried to encourage implicit learning and may have partially succeeded, but there was still no evidence for substantive bias.

CHAPTER 3: POSITIONAL EXTENSION OF AN OBSTRUENT VOICING CONTRAST

The second phonotactic implicational I tested concerns the positional distribution of the voicing contrast in obstruents. 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. Like the implicational about the distribution of major place contrasts, this implicational has a phonetic motivation rooted in perception. Steriade (1997) argued that cues to obstruent voicing are more abundant word-initially than word-finally; in particular, VOT is an available cue word-initially but not word-finally (for [−cont] obstruents[29]). As a result, 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. Thus if languages are phonetically natural, a language that has an obstruent voicing contrast word-finally, where it is harder to perceive, should also have that voicing contrast word-initially, where it is easier to perceive. On the other hand, there is no expectation from phonetic naturalness that a language with an obstruent voicing contrast word-initially should also have that voicing contrast word-finally. In addition to being phonetically motivated, the implicational is supported by the phonological typology (Steriade 1997, Lombardi 1999). Based on data from 52 languages representing 22 families, Blevins (2004:94-95) observes that "[w]ord-final neutralization of laryngeal release features is common, while word-initial neutralization of the same release features is unattested." More specifically, obstruent voicing neutralization tends to occur word-finally. Further evidence for the favoring of voicing contrasts word-initially over word-finally comes from the fact that second language learners tend to master new voicing contrasts in non-final positions better than in final position

---

[29] See Blevins 2004 for discussion of the cues to stop vs. fricative voicing and different phonetic accounts for the development of voicing neutralization in stops and fricatives. I set these points aside since the experiments in this chapter deal only with the distribution of the voicing contrast in stops.

(Broselow 2018). I conducted an experiment in the same vein as the place contrast experiments to test whether the phonotactic implicational about the distribution of the voicing contrast in obstruents is reproduced in artificial grammar learning.

3.1 Experiment 3: Positional Extension of an Obstruent Voicing Contrast with Sonorant Filler Consonants

3.1.1 Method

In Experiment 3, I exposed participants to an obstruent voicing contrast in either word-initial or word-final position and tested whether they extended the contrast to the other position. If participants' behavior is consistent with the phonotactic implicational about voicing contrasts, there should be asymmetric extension: participants exposed to the contrast word-finally, where it is less perceptible, should extend the contrast to word-initial position, where it is more perceptible, more than participants exposed to the contrast word-initially extend it to word-final position. This result would constitute evidence for substantive bias in phonotactic learning.

3.1.1.1 Conditions

While in the place contrast experiments the training patterns varied only in the position in which the contrast occurred, in Experiment 3 the training patterns varied in an additional way. There were four training conditions defined on two dimensions: Trained Contrast Position, that is, whether the language exhibited an obstruent voicing contrast in word-initial or word-final position, and Trained Neutralization Value, that is, whether the language exhibited voiced or voiceless obstruents in the position without a contrast or, put another way, whether obstruents "neutralized" to voiced or voiceless in the position without a contrast. Typologically, neutralization is to the voiceless sound (Steriade 1997, Blevins 2004), but I was interested in whether the voicing value stops “neutralized” to in the position without the contrast affected the degree to which participants extended the voicing contrast to that position. This design allowed me to test for another type of substantive bias related to the markedness of voiced stops (see Section 3.1.2.) The conditions that feature only voiceless stops in word-final position and in word-initial position have phonotactic systems that are consistent with languages that exhibit final devoicing and initial devoicing, respectively. Final devoicing is cross-linguistically prevalent (Lombardi 1991, Blevins 2004), but initial devoicing, while rare, also occurs, in Somali and various dialects of German (Blevins 2004:111, Vietti, Alber, & Vogt 2018). Of course, in accordance with the typological facts discussed above, this initial devoicing does not occur in the absence of final devoicing, so the experimental condition with just initial voiceless stops but both final voiced and voiceless stops still represents an unattested system.

Table 3.1: Experiment 3 Training Conditions

Trained Contrast Position	Trained Neutralization Value	Condition	#T	#D	T#	D#
Final	**Voiced**	**#D…{T, D}# (*#T)**	✗	✓	✓	✓
	Voiceless	***#T…{T, D}# (*#D)***	✓	✗	✓	✓
Initial	Voiced	#{T, D}…D# (*T#)	✓	✓	✗	✓
	Voiceless	*#{T, D}…T# (*D#)*	✓	✓	✓	✗

The properties of the four training conditions are shown in Table 3.1. The condition names express what types of obstruents occurred in which positions in each condition. For instance, in the **#D…{T, D}#** condition, only voiced stops occurred word-initially (e.g. ***b**alin*) while both voiceless and voiced stops occurred word-finally (e.g. *muru**k***, *luna**g***). In the final contrast conditions (bolded), participants heard both voiced and voiceless stops word-finally but only voiced or voiceless stops word-initially. The initial contrast conditions (not bolded) are the mirror image: participants heard both voiced and voiceless stops word-initially but only voiced or voiceless stops word-finally. The conditions in which stops neutralized to voiceless in the position without the contrast are italicized, and the conditions in which stops neutralized to voiced are non-italicized. One can also define the four conditions by what obstruent voicing and position combination is excluded. For example, in the **#D…{T, D}#** condition there are no initial voiceless stops (i.e. *#T). These condition-defining phonotactic constraints are given in parentheses after the condition names in Table 3.1.

As in the place contrast experiments, the test items in Experiment 3 were identical in all conditions. In the test phase, participants in all four conditions heard words beginning and ending with voiced and voiceless stops.

3.1.1.2 Materials

Like in Experiments 1 and 2, the training and test items were all nonce words of the shape $C_1VC_2VC_3$. Either C_1 or C_3 was a stop drawn from [p t k b d ɡ]. The other two Cs were sonorants drawn from [m n l ɹ j w], with the same restrictions on the distribution of the glides [j] and [w]. In each word, all three Cs were different. The vowels were drawn from [i ɑ u]. Voiced and voiceless stops were equally represented in the position with the voicing contrast in training and across positions in test. Additionally, labial, coronal, and velar stops were equally represented across positions and voicing categories in each condition and in both the training and test phases. In the training phase, half of the items with stops in the position with an obstruent voicing contrast were members of a minimal pair for stop voicing (e.g. ***t**uwim-**d**uwim* in *#{T,* *D}…T#* or #{T, D}…D#). This was to encourage participants to notice that voicing was contrastive in stops in the given position. In the test phase, half of all items were members of minimal pairs for voicing. In both training and test, half of the items were iambs and half were trochees. Like in the place contrast experiments, stress did not correlate with the position that featured the obstruent voicing contrast or with the syllable in which the stop (always either word-initial or word-final) occurred. This was to prevent participants from associating the obstruent voicing contrast with stress instead of position and to prevent stress from drawing undue attention to the stops.

Sample training items in the *#{T, D}…T#* condition are shown in Table 3.2. In this condition, participants were trained on words with voiced and voiceless stops in initial position but only voiceless stops in final position. As mentioned above, this training language can be thought of as a final devoicing language, though as this is a phonotactic learning experiment participants were not exposed to any alternations. The full set of training and test stimuli for Experiment 3 are given in the Appendix.

Table 3.2: Sample Training Items in the *#{T, D}…T#* Condition

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

The stimuli were recorded by a phonetically-trained female native speaker of American English. Voiced stops were fully voiced, voiceless stops were aspirated, and word-final stops were released. Unstressed vowels were not reduced. Otherwise, pronunciation was as in American English. The stimuli were produced in isolation as if each item was a sentence unto itself. The stimuli were recorded in a soundproof booth using a Shure SM-10A head-mounted microphone plugged into an XAudioBox. The recordings were done in the program PCQuirerX with a sampling rate of 22,050 Hz.

3.1.1.3 Procedure

The procedure for Experiment 3 was analogous to the procedure for Experiments 1 and 2. Experiment 3 was also conducted online using Experigen, and the instructions to the participants were the same. The experiment began with the same two test words to ensure participants were doing the experiment under adequate listening conditions. The training and test phases proceeded as in Experiments 1 and 2. The training phase consisted of two blocks of the same 36 training items, with images for each item. The test phase consisted of a single block of 48 items, which were the same in all four conditions. After the conclusion of the experiment, participants completed a demographic survey, but in Experiment 3 they did not answer questions about whether they had noticed a pattern in their training language or whether they had used a strategy in the test phase.

Table 3.3: Sample Test Items for Each Training Condition in Experiment 3

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**

As in the place contrast experiments, there were three types of items in the test phase. Familiar Conforming items contained stops whose voicing and position conformed to the trained pattern (e.g. #T, #D, and T# in the *#{T, D}…T#* condition) and were words that were heard in training. Novel Conforming items also contained stops whose voicing and position conformed to the trained pattern, but these words had not been heard in training. Finally, Novel

Nonconforming items featured the voicing and position combination not heard in training (e.g. D# in the *#{T, D}…T#* condition). Table 3.3 gives sample test items for each of the four training conditions.

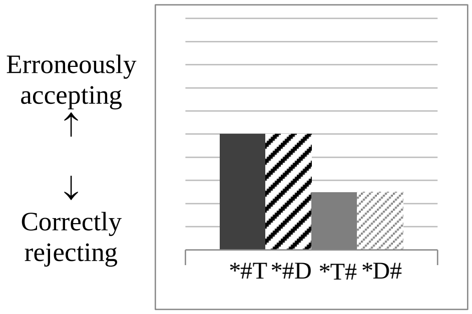
3.1.1.4 Participants

The participants were native English speakers recruited through the UCLA Psychology Subject Pool. None of the participants in Experiment 3 had participated in Experiments 1 or 2. I again excluded participants who were not native English speakers (21 participants), had taken more than one linguistics class (32 participants), reported a history of speech or hearing impairments (4 participants), gave an incorrect response to either of the two test words that preceded the experiment (10 participants), or accepted all test items (4 participants). Additionally, two participants were excluded because the test word sound files did not play properly for them. After exclusions (73 out of 222 participants), there were 149 participants divided among the four conditions. The number of participants per condition ranged from 33 to 41.

3.1.2 Predictions

Because the training conditions in Experiment 3 varied in more ways than the training conditions in the place contrast experiments, the predictions for this experiment are more complex. As in Experiment 1, Familiar Conforming items are predicted to be accepted at a higher rate than Novel Conforming items. As for the rest of the results, different hypotheses make distinct predictions. Participants' 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. For instance, if participants in the *#{T, D}…T#* condition accept test items with word-final voiced stops (D#), they have extended the word-initial obstruent voicing contrast they encountered in training to word-final position.

Figure 3.1: Predicted relative acceptance rates of Novel Nonconforming items by condition according to the position-based substantive bias hypothesis.



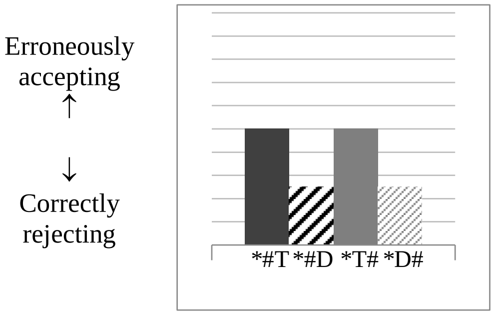
Consider first the position-based substantive bias hypothesis. Recall the phonotactic implicational: an obstruent voicing contrast in final position entails an obstruent voicing contrast in initial position, but not vice versa. Behavior consistent with the implicational would be asymmetric extension: participants exposed to the voicing contrast word-finally should extend it to word-initial position more than participants exposed to the voicing contrast word-initially extend it to word-final position. Extension of the voicing contrast to the other position manifests as erroneously accepting Novel Nonconforming items, so participants trained on the contrast word-finally should more readily accept their Novel Nonconforming items (#T and #D, depending on the condition) than participants trained on the contrast word-initially accept their Novel Nonconforming items (T# and D#, depending on the condition). This pattern of results is illustrated in the schematic graph in Figure 3.1, which shows the predicted relative acceptance rates of Novel Nonconforming items among the four conditions according to the position-based substantive bias hypothesis. Each condition is identified by the type of item that is Novel Nonconforming (i.e. not attested in training) in that condition.

There is a second, voicing-based substantive bias hypothesis. Voiced obstruents are more marked than voiceless obstruents (Greenberg 1978). In particular, voiced stops are harder to produce in word-edge (initial and final) position than voiceless stops (Westbury & Keating 1986). Evidence from second language acquisition also supports the greater markedness of voiced obstruents, as second language learners are better at producing final voiceless obstruents than final voiced obstruents (Broselow 2018). Thus if a language permits voiced stops in a particular word-edge position, it should also permit voiceless stops in that position; not doing so would be phonetically unnatural, from the standpoint of articulatory difficulty. On the other hand, if a language permits voiceless stops in a particular word-edge position, there is no argument from naturalness that it ought to permit voiced stops in that position as well. The voicing-based substantive bias hypothesis therefore predicts more extension from voiced obstruents to voiceless obstruents than from voiceless obstruents to voiced obstruents. That is, participants trained on a neutralizing-to-voiced language should more readily accept their Novel Nonconforming items, which are voiceless (#T or T#), than participants trained on a neutralizing-to-voiceless language accept their Novel Nonconforming items, which are voiced (#D or D#). In other words, there should be higher acceptance rates of Novel Nonconforming items in the conditions where Novel Nonconforming items feature a voiceless stop. This pattern of results is illustrated in the schematic graph in Figure 3.2.[30]

---

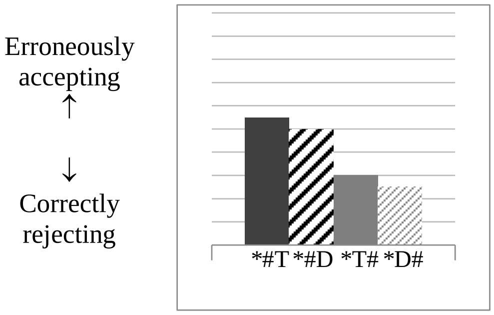
[30] Goldrick (2004) conducted a phonotactic learning study whose results may have shown this type of effect. The experiments were production tasks with speeded repetition, designed to elicit segment replacement errors (e.g. saying /hɛŋ/ for /fɛŋ/), and the purpose of the study was to investigate the level of representation (segmental, featural) at which phonotactic constraints can be encoded in speech production. Participants read words that obeyed certain phonotactic constraints (e.g. /f/ occurs only in onset position (*f#)) but under speeded repetitions conditions sometimes produced errors. Such errors tended to, but did not always, obey the experiment-internal phonotactic restrictions. Goldrick exposed participants to patterns in which both members of a pair of corresponding voiced and voiceless sounds (/s/ and /z/, /f/ and /v/, /t/ and /d/) occurred in one position (word-initial or word-final) but only one member of the voicing pair occurred in the other position. Participants tended to extend more from voiced to voiceless obstruents (e.g. making a speech error that put /s/ in word-final position when they had only been exposed

Figure 3.2: Predicted relative acceptance rates of Novel Nonconforming items by condition according to the voicing-based substantive bias hypothesis.



If both position-based and voicing-based substantive biases are at work, we predict more extension from word-final position to word-initial position as well as more extension from voiced obstruents to voiceless obstruents. One way this pattern of results could manifest is illustrated in the schematic graph in Figure 3.3.

Figure 3.3: Predicted relative acceptance rates of Novel Nonconforming items by condition according to both the position-based and voicing-based substantive bias hypotheses.



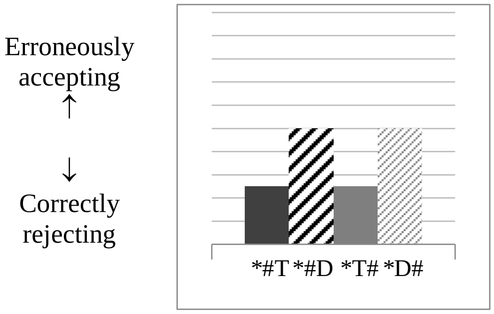
Finally, there is a potential prediction from complexity bias. Due to the presence of sonorant consonants in the training items, the phonotactic constraint that is needed to exclude Novel Nonconforming items in the neutralizing-to-voiceless conditions could be more complex

---

to /z/ word-finally) than from voiceless to voiced obstruents (e.g. making a speech error that put /z/ in word-final position when they had only been exposed to /s/ word-finally). However, these results are only suggestive because Goldrick, for his own purposes, excluded from analysis actual voicing errors (e.g. saying a final /s/ for a final /z/) and because his design, task, and research questions were very different from those of the present study.

than the constraint needed in the neutralizing-to-voiced conditions. To see how this works, we can compare two conditions that differ only in Trained Neutralization Value. In the #{T, D}…D# training condition, a neutralizing-to-voiced condition, participants hear words that end in voiced sonorants (e.g. *kawá**m***) and words that end in voiced obstruents (e.g. *míwi**b***) but no words that end in voiceless obstruents (e.g. *míwi**p***). Thus to exclude the type of item they did not hear, participants could posit the single-feature constraint *[–voice]#. The neutralizing-to-voiceless counterpart of the #{T, D}…D# condition is the *#{T, D}…T#* condition. In this condition, participants hear words that end in voiced sonorants (e.g. *kawá**m***) and words that end in voiceless obstruents (e.g. *míwi**p***) but no words that end in voiced obstruents (e.g. *míwi**b***). To exclude words ending in voiced obstruents while not also excluding words ending in voiced sonorants, participants in this condition must posit the two-feature constraint *[–son, +voice]#. The same is true for the other pair of conditions: participants may learn a one-feature constraint to exclude voiceless obstruents in the neutralizing-to-voiced condition while they must learn a two-feature constraint to exclude voiced obstruents in the corresponding neutralizing-to-voiceless condition. Because it is potentially harder to learn to exclude voiced obstruents, the complexity bias hypothesis predicts more extension from voiceless obstruents to voiced obstruents than from voiced obstruents to voiceless obstruents. Note that this prediction is precisely the opposite of the voicing-based substantive bias prediction described above. The pattern of results predicted by complexity bias are illustrated in the schematic graph in Figure 3.4: Novel Nonconforming items with voiceless stops should be accepted less often (i.e. correctly rejected more often) than Novel Nonconforming items with voiced stops.

Figure 3.4: Predicted relative acceptance rates of Novel Nonconforming items by condition according to the complexity bias hypothesis.



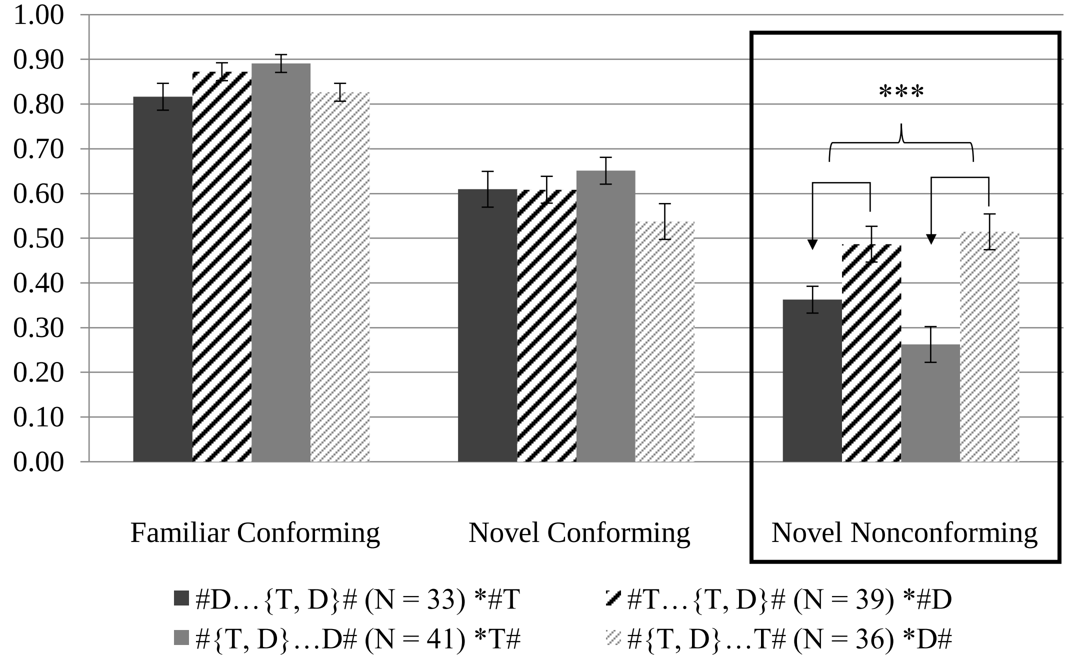
If no synchronic biases are at work, we expect similar acceptance rates of Novel Nonconforming items across all four conditions.

3.1.3 Results

Figure 3.5 shows the acceptance rates of the three types of test items across conditions. Acceptance rates of Novel Nonconforming items are framed and may be informally compared to the schematic graphs in Figures 3.1–3.4. Impressionistically, the pattern of acceptance rates for Novel Nonconforming items most closely remembers the schematic graph in Figure 3.4, which illustrated the effect of complexity bias. This suggests that the results of Experiment 3 are most consistent with complexity bias.

A mixed-effects logistic regression fit to the conforming items (Familiar Conforming and Novel Conforming) with response (accept or reject) as the dependent variable; Familiarity (familiar vs. novel), Trained Contrast Position (word-initial vs. word-final), Trained Neutralization Value (voiced vs. voiceless), and their 2- and 3-way interactions as fixed effects; random intercepts for subject and item; and by-subject random slopes for Familiarity yielded a significant main effect of Familiarity: Familiar Conforming items were accepted more often than Novel Conforming items ($\beta = 1.239$; $p < 0.001$). This is as predicted: participants were better at accepting conforming items they heard in the training phase than conforming items that were new.

Figure 3.5: Acceptance Rate of Test Items by Condition



The acceptance rates of Novel Conforming items were above chance in all four conditions. This was determined by fitting mixed-effects logistic regressions to the novel items (Novel Conforming and Novel Nonconforming) with fixed effects Conformity (conforming vs. nonconforming), Trained Contrast Position, Trained Neutralization Value, and all their 2- and 3-way interactions, random intercepts for subject and item, and by-subject random slopes for Conformity. I changed the reference levels of the factors Trained Contrast Position and Trained Neutralization Value so that in each of the four models the intercept represented the acceptance rate for Novel Conforming items in one of the four conditions. The intercept was significantly above chance in all four models, showing that participants accepted Novel Conforming items at above chance levels in all four conditions.[31] This means that in all conditions participants were able to generalize to novel items that fit their training language, showing that they had learned the pattern they had been trained on. To check that acceptance rates of Novel Conforming items did not differ significantly among the four conditions, I fit a mixed-effects logistic regression just to Novel Conforming items with Condition as the fixed effect and random intercepts for subject and item. There was no significant main effect of Condition, and post-hoc pairwise comparisons (Tukey method) confirmed that no two conditions had significantly different acceptance rates of Novel Conforming items. Consequently, we can compare the acceptance rates of Novel Nonconforming items to test for the different biases reviewed in Section 3.1.2

A mixed-effects logistic regression was fit to the Novel Nonconforming items with response (accept or reject) as the dependent variable, Trained Contrast Position and Trained Neutralization Value as the fixed effects, and random intercepts for subject and item. I did not include the interaction of Trained Contrast Position and Trained Neutralization Value because there was no a priori reason to expect such an interaction. That is, there was no reason to think that the effect of position-based substantive bias might be stronger or weaker depending on neutralization value or that the effect of complexity bias might be stronger or weaker depending on position. Put another way, no theory predicts that the main effects of Trained Contrast Position and Trained Neutralization Value should be anything other than additive. Additionally, I was interested in the overall effect of each factor across both levels of the other factor.[32]

---

[31] The coefficients and p-values for the intercepts representing the acceptance rate of Novel Conforming items in each of the four conditions were as follows: **#D…{T, D}# (*#T)**: $\beta = 0.740$, $p < 0.001$; ***#T…{T, D}# (*#D)***: $\beta = 0.861$, $p < 0.001$; #{T, D}…D# (*T#): $\beta = 0.921$, $p < 0.001$; *#{T, D}…T# (*D#)*: $\beta = 0.504$, $p = 0.015$.

[32] If the interaction of Trained Contrast Position and Trained Neutralization Value is included as a fixed effect in the model, it is not significant ($p = 0.208$). Additionally, I used the R function `anova()` to test whether the interaction contributed significantly to the model, and it did not ($\chi^2(1) = 1.568$, $p = 0.211$).

Recall that the position-based substantive bias hypothesis predicts that participants trained on the obstruent voicing contrast word-finally should accept Novel Nonconforming items more (i.e. extend the contrast to initial position more) than participants trained on the voicing contrast word-initially do. In the regression model, this would manifest as a main effect of Trained Contrast Position such that acceptance rates of Novel Nonconforming items are higher in final contrast conditions than in initial contrast conditions. Complexity bias predicts that participants trained on a neutralizing-to-voiceless language will have a harder time learning to reject Novel Nonconforming items than participants trained on a neutralizing-to-voiced language. In the regression model, this would manifest as a main effect of Trained Neutralization Value such that acceptance rates of Novel Nonconforming items are higher in neutralizing-to-voiceless conditions than in neutralizing-to-voiced conditions. (Note that the voicing-based substantive bias hypothesis predicts the direction of the main effect to be in precisely the opposite direction.)

Table 3.4 gives the fixed effects of the regression model. The main effect of Trained Contrast Position was not significant: acceptance rates of Novel Nonconforming items were not significantly higher in final contrast conditions than in initial contrast conditions. On the other hand, the main effect of Trained Neutralization Value was significant: acceptance rates of Novel Nonconforming items were significantly higher in neutralizing-to-voiceless conditions than in neutralizing-to-voiced conditions.

Table 3.4: Fixed Effects of the Novel Nonconforming Items Regression

	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***

3.1.4 Discussion

The phonotactic implicational I set out to test in Experiment 3 (i.e. an obstruent voicing contrast word-finally entails an obstruent voicing contrast word-initially, but not vice versa) was not reproduced. While within the neutralizing-to-voiced conditions there was more extension from a word-final stop voicing contrast to a word-initial stop voicing contrast than vice versa, this was not the case within the neutralizing-to-voiceless conditions (see Figure 3.5). Thus there was not consistently greater extension of the stop voicing contrast from word-final to word-initial position than from word-initial to word-final position. The substantive bias hypothesis was therefore not supported in this experiment either.

Instead, participants trained on a voicing contrast that neutralized to voiceless stops in one position seemed to extend to voiced obstruents in that position more than participants trained on a voicing contrast that neutralized to voiced stops in one position extended to voiceless obstruents in that position. This is the opposite of the behavior we expect based on the relative markedness of voiced and voiceless stops. Given the presence of (voiced) sonorants in the training items, though, this result can be explained by a complexity bias. In the neutralizing-to-voiced conditions, **#D…{T, D}#** and #{T, D}…D#, participants could respectively learn the one-feature constraints *#[–voice] and *[–voice]# in order to reject Novel Nonconforming items. In contrast, in the neutralizing-to-voiceless conditions, ***#T…{T, D}#*** and *#{T, D}…T#*, participants had to learn the more complex two-feature constraints *#[–son, + voice] and *[–son, + voice]#, respectively, to reject Novel Nonconforming items. While this complexity bias account does provide an explanation for the results of Experiment 3, it depends on English sonorants having an active [+voice] feature even though their voicedness is predictable from their sonorancy. Claiming that English sonorants are [+voice] is not uncontroversial; they have often been treated as underspecified for [voice] instead (see for instance Steriade 1987, Lombardi 1991). That said, some theories that consider sonorants underlyingly underspecified for [voice] still acknowledge that at some level of the grammar this feature is specified as [+voice]. Stemberger (1991) found that, in both a corpus study and an experiment, sonorants interacted more with voiced obstruents than with voiceless obstruents in speech errors. Taking degree of interaction in errors as a measure of similarity, he interprets these results as indicating that sonorants are [+voice] at some point in the derivation since this feature is what makes them more like voiced obstruents than voiceless obstruents. Goldrick (2004) also encounters a case of the nasal /n/ interacting more with /g/ than with /k/ in a speech error experiment. Given this evidence that in production sonorants are treated as more similar to voiced obstruents than to voiceless ones, the complexity bias account of the results of Experiment 3 seems plausible. Even if English sonorants are underlyingly underspecified for [voice], the level at which participants are forming generalizations may be a level at which sonorants are [+voice].

Moreover, this surprising result whereby the phonotactic equivalent of a final voicing language is learned better than the phonotactic equivalent of a final devoicing language is not unprecedented. The unexpected results of Greenwood's (2016) phonotactic learning experiment exhibit a similar pattern. Greenwood's experiment had two conditions: in one condition, participants were exposed to a language with a phonotactic restriction against word-final voiceless obstruents, and in the other participants were exposed to a language with a phonotactic restriction against word-final voiced obstruents. In Greenwood's experiments, obstruents were fricatives and affricates rather than stops, so her experiments' phonotactic constraints can be summarized as *S# and *Z#. Crucially, as in my experiment, the training items in both of Greenwood's conditions also included words that ended in (voiced) sonorants. (Unlike my experiment, Greenwood's experiment also included training items ending in vowels.) Thus in her experiment participants in the *S# condition could have posited the one-feature constraint *[–voice]# to correctly exclude words ending in voiceless obstruents while participants in the *Z# condition would have had to posit the two-feature constraint *[–son, +voice]# to correctly exclude words ending in voiced obstruents but not words ending in voiced sonorants (or vowels). The situation was analogous to the situation in Experiment 3.

Greenwood tested the learnability of her two languages in a casual speech condition and a careful speech condition. In the casual speech condition, where Greenwood expected the more natural *Z# language to be learned better, there was no difference in how well the two languages were learned. In the careful speech condition, where Greenwood expected the two languages to be learned equally well, the *S# (final voicing) language was learned better than the *Z# (final devoicing) language. Given that the stimuli in all my experiment were produced carefully, Experiment 3 more closely resembles Greenwood's careful speech condition, and our results are analogous. While Greenwood does not consider complexity bias as a possible explanation for her unexpected results, in both our experiments participants learned the training pattern better when they could posit the constraint *[–voice]# than when they had to posit the more complex constraint *[–son, +voice]#. Complexity bias can explain the unusual results of both of our studies.

Interestingly, Myers & Padgett's (2014) Experiment 1 is quite similar to Greenwood's experiment and the initial contrast conditions of Experiment 3 but did not turn up a similar result. Myers & Padgett trained participants on either a phonotactic restriction against phrase-final voiceless obstruents or a phonotactic restriction against phrase-final voiced obstruents. In fact, the obstruents in question were just [s] and [z], so the restrictions can be expressed with the constraints *s% and *z%, where % represents a phrase boundary. Like the patterns in Greenwood's and my experiments, these patterns are the phonotactic equivalents of final voicing and final devoicing, respectively. Also like in Greenwood's and my experiments, both training languages featured phrases ending in voiced sonorants (in this case, just nasals). (Similarly to Greenwood, Myers & Padgett also included phrases ending in vowels.) Thus the same complexity prediction could be made for Myers & Padgett's experiment: the *s% pattern should be learned better because participants can learn the constraint *[–voice]% to exclude phrase-final [s]s while for the *z% pattern participants must learn the constraint *[–son, +voice]% to exclude phrase-final [z]s without also excluding phrase-final sonorants. This was not Myers & Padgett's result, however. They do not demonstrate that the phrase-final patterns were mastered to different degrees, and in fact the natural but complex *z% pattern appears to have been learned slightly, if not significantly, better than the *s% pattern. Greenwood (2016) also noticed the difference between her result and Myers & Padgett's. The fact that Myers & Padgett did not find the same effect that Greenwood and I found may have to do with the fact that in Myers & Padgett's experiment participants were only exposed to a single licit obstruent ([s] or [z]) in each condition. Perhaps learning that exactly one obstruent was allowed phrase-finally was accomplished through very explicit, even non-linguistic, learning.

Experiment 3 joins Experiments 1 and 2 in turning up no evidence for substantive bias in phonotactic learning. Neither of the two phonotactic implicationals about the positional distribution of contrasts was reproduced in artificial grammar learning. Experiments 1 and 2 showed that a language with major place contrasts only word-initially and a language with major place contrasts only word-finally were equally learnable. Experiment 3 showed that languages with an obstruent voicing contrast only word-finally were as learnable as languages with an obstruent voicing contrast only word-initially. Unfortunately, participants in Experiment 3 were not asked whether they had noticed patterns or what strategies they had used in the test phase, so it is not possible to gauge how explicit learning was in this experiment. While the different training languages in Experiments 1 and 2 were learned equally well, the training languages in Experiment 3 were not all learned to an equal degree. The differences in their learnability, however, point only to an effect of complexity bias, not to an effect of substantive bias.

3.2 Experiment 4: Positional Extension of an Obstruent Voicing Contrast with Voiceless Fricative Filler Consonants

The results of Experiment 3 provided no support for substantive bias, but I argued that they did suggest an effect of complexity bias. The language prohibiting final voiceless stops (*T#) was learned significantly better than the language prohibiting final voiced stops (*D#). Similarly, the language prohibiting initial voiceless stops (*#T) was learned better than the language prohibiting initial voiced stops (*#D), though the difference did not reach significance. These preferences are the opposite of those one would predict based on the relative markedness of voiced and voiceless stops. Since voiced stops are more marked than voiceless stops, one would expect voiced stops to be easier to exclude than voiceless stops. Instead, the opposite pattern emerged, and I proposed that this was due to a complexity bias that depended on the broader phonological structure of the training languages. All the non-critical consonants in the stimuli in Experiment 3 were voiced sonorants, so if learners could group voiced stops and voiced sonorants together as [+voice], then to exclude voiceless stops they only needed to learn the constraint *[–voice], for either word-initial or word-final position depending on their training language. If voiced stops and voiced sonorants are both [+voice], though, then to exclude voiced stops learners had to learn the constraint *[–son, +voice] for either word-initial or word-final position. This constraint is more complex by virtue of containing an additional feature and may therefore have been relatively more difficult to learn, leading to poorer performance in the conditions prohibiting voiced stops than in the conditions prohibiting voiceless stops.

The phonotactic constraints that could or had to be used to successfully reject nonconforming items in Experiment 3 depended crucially on the non-critical consonants in the training language, namely, the voiced sonorants. If the non-critical consonants were something else, the phonotactic constraints learners could use or would need to use would also change. Additionally, my complexity bias interpretation of the results of Experiment 3 depends on English sonorants being truly [+voice], as opposed to underspecified for voicing. To test the validity of my interpretation of Experiment 3 and to investigate complexity bias further, I conducted another experiment modeled on Experiment 3 but designed to reverse the pattern of results predicted by complexity bias.

The design of Experiment 4 was identical to that of Experiment 3 except that the non-critical consonants in the stimuli were voiceless fricatives instead of voiced sonorants. There were still two conditions in which voiced stops were absent at one edge of the word and two conditions in which voiceless stops were absent at one edge of the word. Since the non-critical consonants were all voiceless, though, now voiced stops were easier to exclude than voiceless stops. To correctly reject voiced stops in a given position, learners could simply learn the constraint *[+voice] in that position. On the other hand, to correctly reject voiceless stops in a given position without also incorrectly rejecting licit voiceless fricatives, learners had to learn the more complex constraint *[–cont, –voice] in that position. Essentially, by changing all the non-critical consonants from voiced sonorants to voiceless fricatives, the prediction from complexity bias is flipped relative to Experiment 3: in Experiment 4, learners should have more trouble learning to exclude *voiceless* stops than voiced stops. If the results of Experiment 4 are the mirror image of the results of Experiment 3, this would support the complexity bias interpretation for Experiment 3 and provide further evidence that complexity bias plays a role in phonotactic learning.

3.2.1 Method

Experiment 4 is analogous to Experiment 3. I exposed participants to a stop voicing contrast in word-initial or word-final position and tested whether they extended it to the other position. Experiment 4 still tested whether learners would reproduce the phonotactic implicational whereby an obstruent voicing contrast in word-final position entails an obstruent voicing contrast in word-initial position but not vice versa. However, Experiment 4 was also specifically designed to test the correctness of my complexity bias interpretation of the results of Experiment 3.

3.2.1.1 Conditions

The four conditions in Experiment 4 were the same as the four conditions in Experiment 3 (see Table 3.1). The training languages featured a stop voicing contrast in either word-initial or word-final position. In the position without the stop voicing contrast, stops were either only voiceless or only voiced. In the test phase, participants in all four conditions were asked to accept or reject words beginning and ending with voiced and voiceless stops.

3.2.1.2 Materials

The stimuli in Experiment 4 were analogous to the stimuli in Experiment 3. As in Experiment 3, the training and test items were nonce words of the shape $C_1VC_2VC_3$, and either C_1 or C_3 was a stop drawn from [p t k b d ɡ]. The other two Cs, however, instead of being sonorants, were voiceless fricatives drawn from [f θ s ʃ h]. The fricative [h] was restricted to word-initial position. In order to make Experiment 4 differ from Experiment 3 as minimally as possible, I created the stimuli for Experiment 4 by taking the stimuli from Experiment 3 and systematically converting the sonorants in those stimuli to voiceless fricatives. I did this by establishing a correspondence between the sonorants used in Experiment 3 and the voiceless fricatives used in Experiment 4. I replaced each instance of [m], [n], [ɹ], and [l] with [f], [θ], [s], and [ʃ], respectively. In Experiment 3, [m n ɹ l] were the only sonorants to occur word-finally since the glides [j] and [w] were excluded from that position, so the sonorant-voiceless fricative correspondence was one-to-one in word-final position. I replaced word-initial instances of [j] with [h]. This left word-initial [w]s without a unique voiceless fricative correspondent, so I converted these [w]s to [f], [s], or [h]. Word-medially, both [j] and [w] were left without unique voiceless fricative correspondents, so I converted these glides to [f], [s], or [ʃ]. I did not reassign any glides to [θ] because this fricative is relatively infrequent in English. I did not want to include any more [θ]s than necessary lest the training languages sound even less English-like than they already did.

Apart from the replacement of the sonorants with voiceless fricatives, the stimuli were identical to the stimuli in Experiment 3. In each word, all three Cs were different, and the vowels were drawn from [i ɑ u]. The distribution of voiced and voiceless stops, of labial, coronal, and velar stops, and of trochees and iambs was the same, as was the structure of the training and test items in terms of minimal pair membership.

Sample training items in the *#{T, D}…T#* condition of Experiment 4 are shown in Table 3.5. The training language in this condition is the "final devoicing" language: participants were trained on words with voiced and voiceless stops in initial position but only voiceless stops in final position. The full set of training and test stimuli for Experiment 4 are given in the Appendix.

Table 3.5: Sample Training Items in the *#{T, D}…T#* Condition

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

The stimuli were recorded by the same female speaker who recorded the stimuli for Experiment 3. As in the other experiments, voiced stops were fully voiced, voiceless stops were aspirated, word-final stops were released, unstressed vowels were not reduced, and the stimuli were produced in isolation as if each item was a sentence unto itself. The stimuli were recorded in a soundproof booth using a Shure SM-10A head-mounted microphone plugged into an audio box. The recordings were done in the program PCQuirerX with a sampling rate of 22,050 Hz.

3.2.1.3 Procedure

The procedure for Experiment 4 was identical to the procedure for Experiment 3 except that in Experiment 4, participants did answer questions about whether they had noticed a pattern in their training language or whether they had used a strategy in the test phase. Experiment 4 is like Experiments 1 and 2 and unlike Experiment 3 in asking participants how they approached the task in the post-experiment survey.

As in Experiment 3, the test phase featured Familiar Conforming, Novel Conforming, and Novel Nonconforming items. Table 3.6 gives sample test items for each of the four training conditions in Experiment 4.

Table 3.6: Sample Test Items for Each Training Condition in Experiment 4

Condition	**Familiar Conforming**	**Novel Conforming**	**Novel Nonconforming**
#D…{T, D}# (*#T)	θifá**p**	síθu**p**	**p**ísuf
#T…{T, D}# (*#D)	θifá**p**	síθu**p**	**b**ísuf
#{T, D}…D# (*T#)	**k**aʃáf	**k**áfis	fúʃi**k**
*#{T, D}…T# (*D#)*	**k**aʃáf	**k**áfis	fúʃi**g**

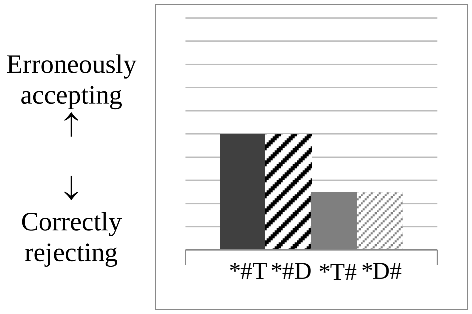
3.2.1.4 Participants

The participants were native English speakers recruited through the UCLA Psychology Subject Pool. To the best of my knowledge, none of the participants in Experiment 4 had participated in Experiments 1–3. I applied the same exclusion criteria as in Experiments 1–3 and excluded 96 out of 240 participants. Of the 96 excluded participants, 30 were not in fact native English speakers, 28 had taken more than one linguistics class, 17 reported a history of speech or hearing impairments, 14 gave an incorrect response to either of the two test words that preceded the experiment, 4 accepted all test items, and 3 experienced technical problems with the experiment that required their data to be thrown out. After exclusions, there were 144 participants divided among the four conditions.

3.2.2 Predictions

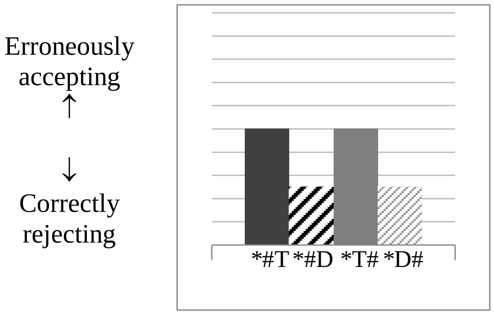
Because the design of Experiment 4 is analogous to that of Experiment 3, the same types of predictions exist for Experiment 4. In terms of substantive bias, the predictions are the same as in Experiment 3. The position-based substantive bias hypothesis predicts more extension of the stop voicing contrast from word-final to word-initial position than from word-initial to word-final position. This would manifest as higher acceptance rates of Novel Nonconforming items in the conditions in which participants were trained on the contrast word-finally than in the conditions in which participants were trained on the contrast word-initially. This pattern of results was schematized in Figure 3.1, repeated below as Figure 3.6.

Figure 3.6: Predicted relative acceptance rates of Novel Nonconforming items by condition according to the position-based substantive bias hypothesis.



The voicing-based substantive bias hypothesis predicts more extension from voiced stops to voiceless stops than from voiceless stops to voiced stops. This would manifest as higher acceptance rates of Novel Nonconforming items in the neutralizing-to-voiced conditions than in the neutralizing-to-voiceless conditions. That is, acceptance rates of Novel Nonconforming items should be higher in the conditions where Novel Nonconforming items contain voiceless stops than in the conditions where Novel Nonconforming items contain voiced stops. This pattern of results was schematized in Figure 3.2, repeated below as Figure 3.7.

Figure 3.7: Predicted relative acceptance rates of Novel Nonconforming items by condition according to the voicing-based substantive bias hypothesis or the complexity bias hypothesis.

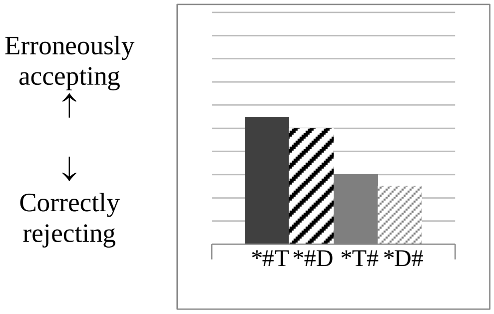


It is in the prediction from complexity bias that Experiment 4 differs from Experiment 3. By changing the voicing of the non-critical consonants in the artificial languages, I have reversed the prediction. The training languages that could be mastered with a simpler constraint in Experiment 3, that is, the neutralizing-to-voiced languages, require a more complex constraint in Experiment 4 while the training languages that required a more complex constraint in Experiment 3, that is, the neutralizing-to-voiceless languages, can be mastered with a simpler constraint in Experiment 4. The situation in Experiment 4 can be exemplified with the two word-initial contrast conditions. In the *#{T, D}…T#* training condition, a neutralizing-to-voiceless condition, participants hear words that end in voiceless fricatives (e.g. *kaʃá**f***) and words that end in voiceless stops (e.g. *físi**p***) but no words that end in voiced stops (e.g. *físi**b***). Thus to exclude the type of item they did not hear, participants could posit the single-feature constraint *[+voice]#. The neutralizing-to-voiced counterpart of the *#{T, D}…T#* condition is the #{T, D}…D# condition. In this condition, participants hear words that end in voiceless fricatives (e.g. *kaʃá**f***) and words that end in voiced stops (e.g. *físi**b***) but no words that end in voiceless stops (e.g. *físi**p***). To exclude words ending in voiceless stops while not also excluding words ending in voiceless fricatives, participants in this condition must posit the two-feature constraint *[–cont, –voice]#. The same is true for the word-final contrast conditions: participants may learn a one-feature constraint to exclude voiced stops in the neutralizing-to-voiceless condition while they must learn a two-feature constraint to exclude voiceless stops in the corresponding neutralizing-to-voiced condition. Because it is potentially harder to learn to exclude voiceless stops, the complexity bias hypothesis predicts more extension from voiced stops to voiceless stops than from voiceless stops to voiced stops.

Note that this is precisely the same prediction the voicing-based substantive bias hypothesis makes. In Experiment 3, the voicing-based substantive bias hypothesis and the complexity bias hypothesis made directly opposite predictions, but in Experiment 4, due to the type of non-critical consonant and the resulting phonological structure of the training languages, the prediction from complexity bias aligns with the prediction based on the relative markedness of voiced and voiceless stops. The pattern of results that would be consistent with complexity bias is also schematized in Figure 3.7, the same figure that schematizes the pattern of results that would be consistent with the voicing-based substantive bias. By virtue of its design, Experiment 4 cannot disambiguate the voicing-based substantive bias and complexity bias. However, if the results of Experiment 4 fit both of these hypotheses, I can argue that the effect is one of complexity bias and not of substantive bias by appealing to the results of Experiment 3, where these two biases made opposing predictions. Experiment 3 provided no support for the voicing-based substantive bias, and since Experiment 4 is modeled very closely on Experiment 3, I do not expect any effect of voicing-based substantive bias here either. Thus if the results match the predictions made by both the voicing-based substantive bias hypothesis and the complexity bias hypothesis, I will be inclined to attribute them to complexity bias.

Finally, there is the possibility that the results could show the effects of multiple biases. As in Experiment 3, the position-based and voicing-based substantive biases in combination would result in more extension from word-final position to word-initial position and more extension from voiced obstruents to voiceless obstruents. One potential manifestation of this pattern of results was schematized in Figure 3.3, repeated below as Figure 3.8.

Figure 3.8: Predicted relative acceptance rates of Novel Nonconforming items by condition according to both the position-based and voicing-based substantive bias hypotheses or both the position-based substantive bias and the complexity bias hypotheses.

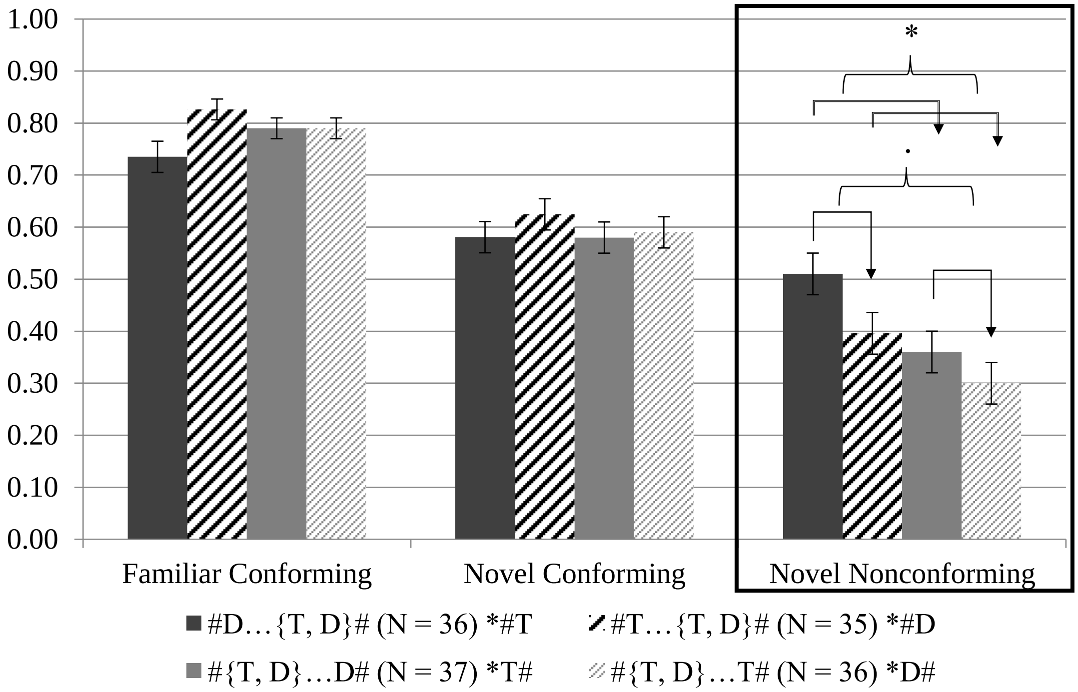


Recall that in Experiment 4 the complexity bias hypothesis also predicts more extension from voiced obstruents to voiceless obstruents. Thus the schematic pattern shown in Figure 3.8 could also reflect a combination of the effects of position-based substantive bias and complexity bias. Again, in Experiment 4 an effect of voicing-based substantive bias cannot be distinguished from an effect of complexity bias, but since Experiment 3 could distinguish these effects and showed an effect of complexity bias, not voicing-based substantive bias, if the relevant effect emerges in Experiment 4 I will attribute it to complexity bias too.

3.2.3 Results

Figure 3.9 shows the acceptance rates of the three types of test items across conditions. Acceptance rates of Novel Nonconforming items are framed and may be informally compared to the schematic graphs in Figures 3.6–3.8. The pattern of acceptance rates of Novel Nonconforming items most closely resembles the schematic graph in Figure 3.8, which illustrates the combined effects of position-based substantive bias and complexity bias. Thus it appears that in Experiment 4 both learning biases were active.

Figure 3.9: Acceptance Rates of Test Items by Condition



A mixed-effects logistic regression fit to the conforming items (Familiar Conforming and Novel Conforming) with response (accept or reject) as the dependent variable; Familiarity (familiar vs. novel), Trained Contrast Position (word-initial vs. word-final), Trained Neutralization Value (voiced vs. voiceless), and their 2- and 3-way interactions as fixed effects; random intercepts for subject and item; and by-subject random slopes for Familiarity yielded a significant main effect of Familiarity: Familiar Conforming items were accepted more often than Novel Conforming items ($\beta = 0.803$; $p < 0.001$). Like in Experiments 1 and 3, and as predicted, participants were better at accepting conforming items they had heard before than conforming items they had not heard before.

Like in Experiment 3, the acceptance rates of Novel Conforming items were above chance in all four conditions. This was determined by the same method as for Experiment 3.[33] Thus in all conditions participants could generalize to novel items that fit their training language, demonstrating that they had learned the pattern they had been trained on. To check that acceptance rates of Novel Conforming items did not differ significantly among the four conditions, I fit a mixed-effects logistic regression just to Novel Conforming items with Condition as the fixed effect and random intercepts for subject and item. There was no significant main effect of Condition, and post-hoc pairwise comparisons (Tukey method) confirmed that no two conditions had significantly different acceptance rates of Novel Conforming items. As for Experiment 3, then, we can compare the acceptance rates of Novel Nonconforming items to test for different biases.

As mentioned above, the framed Novel Nonconforming acceptance rates in Figure 3.9 most closely resemble the schematic graph in Figure 3.8, which illustrated the effect of position-based substantive bias combined with the effect of either voicing-based substantive bias or complexity bias. Just in terms of mean acceptance rates, within both neutralizing-to-voiced conditions and neutralizing-to-voiceless conditions, participants accepted novel nonconforming items, i.e. extended the voicing contrast, more when they had been trained on the contrast word-finally than when they had been trained on the contrast word-initially. The direction of these differences is consistent with position-based substantive bias. Additionally, within both word-final contrast conditions and word-initial contrast conditions, participants extended more from voiced stops to voiceless stops than vice versa. The direction of these differences is consistent

---

[33] The coefficients and *p*-values for the intercepts representing the acceptance rate of Novel Conforming items in each of the four conditions were as follows: **#D…{T, D}# (*#T)**: $\beta = 0.488$, $p = 0.005$; ***#T…{T, D}# (*#D)***: $\beta = 0.735$, $p < 0.001$; #{T, D}…D# (*T#): $\beta = 0.484$, $p = 0.005$; *#{T, D}…T# (*D#)*: $\beta = 0.482$, $p = 0.006$.

with complexity bias. (It is also consistent with voicing-based substantive bias, but as there was no evidence for this bias in Experiment 3 and there was evidence for complexity bias, I opt for the complexity bias explanation here.)

To test the significance of these differences, I fit a mixed-effects logistic regression to the Novel Nonconforming items with response (accept or reject) as the dependent variable, Trained Contrast Position and Trained Neutralization Value as the fixed effects, and random intercepts for subject and item. I again did not include the interaction of Trained Contrast Position and Trained Neutralization Value.[34] The position-based substantive bias hypothesis predicts that participants trained on the stop voicing contrast word-finally should accept Novel Nonconforming items more (i.e. extend the contrast to initial position more) than participants trained on the voicing contrast word-initially do. This would manifest as a main effect of Trained Contrast Position such that acceptance rates of Novel Nonconforming items are higher in final contrast conditions than in initial contrast conditions. In Experiment 4, in which non-critical consonants were voiceless fricatives, complexity bias predicts that participants trained on a neutralizing-to-voiced language will have a harder time learning to reject Novel Nonconforming items (i.e. voiceless stops) than participants trained on a neutralizing-to-voiceless language. This would manifest as a main effect of Trained Neutralization Value such that acceptance rates of Novel Nonconforming items are higher in neutralizing-to-voiced conditions than in neutralizing-to-voiceless conditions.

The fixed effects of the regression model are given in Table 3.7. Unlike in Experiment 3, the main effect of Trained Contrast Position was significant: acceptance rates of Novel

---

[34] If the interaction of Trained Contrast Position and Trained Neutralization Value is included as a fixed effect in the model, it is not significant ($p = 0.699$). Additionally, I used the R function `anova()` to test whether the interaction contributed significantly to the model, and it did not ($\chi^2(1) = 0.148$, $p = 0.701$).

Nonconforming items were higher in final contrast conditions than in initial contrast conditions. There was also a main effect of Trained Neutralization Value: acceptance rates of Novel Nonconforming items were higher in neutralizing-to-voiced conditions than in neutralizing-to-voiceless conditions. While the main effect supporting complexity bias was significant in Experiment 3, in Experiment 4 it was only marginally significant.

Table 3.7: Fixed Effects of the Novel Nonconforming Items Regression

	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

3.2.4 Discussion

In Experiment 4, unlike in Experiment 3, participants did behave in accordance with the phonotactic implicational about the positional distribution of the obstruent voicing contrast. There was greater extension of the voicing contrast from word-final to word-initial position than from word-initial to word-final position, supporting the position-based substantive bias hypothesis.

Additionally, participants trained to neutralize to voiced stops in one position extended to voiceless stops in that position more than participants trained to neutralize to voiceless stops in one position extended to voiced stops in that position. That is, the acceptance rates of Novel Nonconforming items were higher in conditions in which participants had to learn the constraint *[–cont, –voice] to exclude nonconforming items than in conditions in which participants could learn the constraint *[+voice] to exclude nonconforming items. This effect is consistent with complexity bias. It is also consistent with voicing-based substantive bias, but as there was no evidence for this bias in Experiment 3, I attribute the effect in Experiment 4 to complexity bias.

The main effect of Trained Neutralization Value was only marginally significant, suggesting that the effect of complexity bias was weaker in Experiment 4 than in Experiment 3.

The results of Experiment 4 do make a striking point about the role of non-critical consonants and the overall structure of the artificial language. Experiments 3 and 4 both tested the learnability of phonotactic systems with different distributions of voiced and voiceless stops. The experiments had identical structures, and the stimuli in the two experiments were kept as similar as possible, the crucial difference being that the non-critical (i.e. non-stop) consonants in Experiment 3 were voiced sonorants while in Experiment 4 they were voiceless fricatives. That is, in Experiment 3 the non-critical consonants shared their voicing value with one set of critical consonants, and in Experiment 4 they shared their voicing value with the other set of critical consonants.

One might have imagined that filler consonants would be inert within the experiment and that participants would either be equally willing to extend to voiced or voiceless stops in the position in which they had not heard a stop voicing contrast or would exhibit asymmetric extension rooted in the greater markedness of voiced stops relative to voiceless stops. With Experiments 3 and 4, I have shown that this is not the case. Instead, participants take into account the features of all segments represented in their training languages and seem able to infer phonotactic constraints accordingly, opting for the simplest constraint necessary to master their language.

Although the main effect of Trained Neutralization Value did not quite reach significance in Experiment 4, the change in the direction of this main effect from Experiment 3 to Experiment 4 is revealing. In Experiment 3, where non-critical consonants were voiced, participants were always better at correctly rejecting Novel Nonconforming items in the condition in which Novel

Nonconforming items featured a voiceless stop than in the corresponding condition in which Novel Nonconforming items featured a voiced stop. In Experiment 4, on the other hand, where non-critical consonants were voiceless, participants were always better at correctly rejecting Novel Nonconforming items in the condition in which Novel Nonconforming items featured a voiced stop than in the corresponding condition in which Novel Nonconforming items featured a voiceless stop. Participants were therefore always better at rejecting Novel Nonconforming items featuring stops whose voicing feature was the opposite of the voicing feature of the non-critical consonants. This is because the non-critical consonants, and therefore their voicing feature, occur in all positions, thus providing evidence that segments with that voicing feature can occur in all positions and, more importantly, in the position in which a particular type of stop (voiced or voiceless) is absent. When the non-critical consonants' voicing feature matches the voicing feature of the type of stop that is unattested in a given position (e.g. voiceless fricatives are [–voice] and voiceless stops are unattested word-finally), it is harder to correctly reject the unattested stop type. When the non-critical consonants' voicing feature is the opposite of the voicing feature of the type of stop that is unattested in a given position (e.g. voiceless fricatives are [–voice] and voiced stops are unattested word-finally), it is easy to reject the unattested stop type since the non-critical consonants are not providing conflicting evidence about the permissibility of a particular value of the voicing feature in a particular position. These results suggest that in Experiment 3 (voiced) sonorants and voiced stops are being opposed with voiceless stops (see Section 3.1.4 for discussion of the voicing feature of English sonorants) while in Experiment 4 voiceless fricatives and voiceless stops are being opposed with voiced stops.

Experiments 3 and 4 together demonstrate that the experiment-internal distribution of sounds, including filler or non-critical sounds, can have a profound impact on performance. The direction of the differences in acceptance rates of Novel Nonconforming items between neutralizing-to-voiced and neutralizing-to-voiceless conditions flipped from one experiment to the other simply due to the voicing of the non-critical consonants. This effect crucially depended on all the non-critical consonants having the same voicing within a given experiment. I predict that an experiment modeled on Experiments 3 and 4 but with both voiced and voiceless filler consonants (e.g. sonorants and voiceless fricatives) should yield no difference in the acceptance rates of Novel Nonconforming items between corresponding neutralizing-to-voiced and neutralizing-to-voiceless conditions. In such an experiment, it would no longer be the case that stops with one voicing value could be disallowed with a one-feature constraint while stops with the other voicing value could only be disallowed with a two-feature constraint because the filler consonants would no longer all align with one voicing value or the other.

As noted above, the main effect of Trained Neutralization Value, which supports complexity bias, is weaker in Experiment 4 than in Experiment 3. This difference between the experiments actually provides further evidence against a voicing-based substantive bias being active in phonotactic learning. Recall that in Experiment 4, in which non-critical consonants were voiceless fricatives, the predictions from voicing-based substantive bias and complexity bias were identical: participants should accept Novel Nonconforming items less in neutralizing-to-voiceless languages than in neutralizing-to-voiced languages. Such a result could mean that voiced stops were easier to reject than voiceless stops either because they are more marked or because they could be excluded by learning the constraint *[+voice] while excluding voiceless stops, but not voiceless fricatives, required learning the constraint *[–cont, –voice]. While

Experiment 3 provided no evidence for the voicing-based substantive bias, it might still have emerged in Experiment 4. Moreover, it was possible that voicing-based substantive bias and complexity bias could have both had an effect on acceptance rates of Novel Nonconforming items in Experiment 4. That is, if Novel Nonconforming items featuring voiced stops were easier to reject both because voiced stops are more marked than voiceless stops and because they could be excluded with a simpler constraint, then there should have been a compounding effect of voicing-based substantive bias and complexity bias resulting in even lower acceptance rates of Novel Nonconforming items in neutralizing-to-voiceless conditions in Experiment 4. In Experiment 3, voicing-based substantive bias and complexity bias made opposite predictions, but in Experiment 4, they could jointly result in a more extreme effect of complexity bias that might in fact represent the combined effects of two types of bias. However, the apparent effect of complexity bias in Experiment 4 is not more extreme but rather *less* extreme than in Experiment 3. The differences in acceptance rates of Novel Nonconforming items between neutralizing-to-voiced and neutralizing-to-voiceless conditions are smaller, not larger. This is the opposite of what we expect if there is a compounding effect of voicing-based substantive bias and complexity bias in Experiment 4, suggesting that voicing-based substantive bias is not active, or at the very least not additive in its effect.

The weaker effect of Trained Neutralization Value in Experiment 4 relative to Experiment 3 supports my claim that voicing-based substantive bias does not play a role in these results. The question of why this effect is weaker in Experiment 4 remains, though. Experiments 3 and 4 are mirror images of each other with respect to the voicing of the non-critical consonants and therefore with respect to the predictions of complexity bias, so why do the complexity bias effects fall short of significance in Experiment 4? The logical place to look for an explanation is in the difference between sonorants and fricatives as non-critical consonants. In an ideal world, the non-critical consonants in Experiments 3 and 4 would differ *only* in voicing, but given that I constructed stimuli using the phoneme inventory of English, this was not possible. Consequently, there was another fundamental difference between the non-critical consonants in each experiment: in Experiment 3, the non-critical consonants were necessarily voiced by virtue of their being English sonorants while in Experiment 4, the non-critical consonants were voiceless by design, not by virtue of their being English fricatives. English of course has voiced fricatives as well as voiceless fricatives, but it does not have voiceless sonorants. This means that in Experiment 3, as long as the native English-speaking participants perceived the non-critical consonants as sonorants, they knew they were voiced. Native English speakers would not misperceive a sonorant as being voiceless, so in the neutralizing-to-voiced conditions, there was no chance that participants might think they had heard a voiceless segment in the position in which only voiced segments (sonorants or voiced stops) occurred.

In Experiment 4, in contrast, the non-critical consonants were voiceless fricatives and could have been misperceived as devoiced voiced fricatives since English does in fact have voiced fricatives and also sometimes exhibits devoicing of voiced obstruents, including fricatives (Lisker & Abramson 1967, Haggard 1978, Stevens et al. 1992, Smith 1997). In other words, in some cases participants may have mistakenly “undone” devoicing and taken voiceless fricatives to be underlyingly voiced fricatives (i.e. they engaged in hypercorrection (Ohala 1992)). If some voiceless fricatives in the neutralizing-to-voiceless conditions were miscategorized as voiced, then it would no longer be the case that unattested voiced stops could be excluded with the one-feature constraint *[+voice], since such a constraint would also bar voiced fricatives, which participants mistakenly believed were present and allowed in the position without the stop voicing contrast. Thus participants would have had to posit *[–cont, +voice], making their task no different than in the neutralizing-to-voiced conditions, where participants had to posit *[–cont, –voice].

In sum, because the participants' native language contains voiced fricatives but does not contain voiceless sonorants, it was possible for participants to misperceive the voicing of non-critical consonants in Experiment 4 but not in Experiment 3. Misperception in Experiment 4 could have prevented participants from being able to learn a one-feature constraint in the neutralizing-to-voiceless conditions, making those conditions just as complex as the neutralizing-to-voiced conditions and thus reducing the difference in acceptance rates of Novel Nonconforming items between conditions differing in Trained Neutralization Value. This would explain why the main effect predicted by complexity bias is less robust in Experiment 4 than in Experiment 3. If this account is correct, it would also demonstrate that the structure of the phonemic inventory of the participants' native language, as well as the structure of the intra-experimental segment inventory, can influence performance in artificial grammar learning experiments.

3.3 Combined Analysis of Experiments 3 and 4

Experiments 3 and 4 are nearly identical studies; they differ only in the nature of the non-critical consonants in their stimuli. Both were designed to test whether learners would reproduce the phonotactic implicational whereby an obstruent voicing contrast in word-final position entails one in word-initial position but not the other way around. The four conditions in each experiment were structured in the same way such that comparing two pairs of conditions could test for an effect of substantive bias and comparing two pairs of conditions could test for an effect of complexity bias. This similarity in structure makes it possible to pool the results of the two experiments and carry out a combined analysis with more power.

In combining the results of Experiments 3 and 4, I pair the conditions in each experiment that match in expected naturalness and in expected complexity, given the identity of the non-critical consonants. Expected naturalness follows from Trained Contrast Position. The word-initial contrast training languages are more natural insofar as it is natural to have an obstruent voicing contrast word-initially but not word-finally. Participants in the word-initial contrast conditions are expected to accept Novel Nonconforming items less, meaning that they extend the initial voicing contrast to final position less and have therefore learned their training language better. The word-final contrast training languages are less natural insofar as it is unnatural to have an obstruent voicing contrast word-finally but not word-initially. Participants in the word-final contrast conditions are expected to accept Novel Nonconforming items more, meaning that they extend the final voicing contrast to initial position more and have therefore not learned their training language as well. In both Experiments 3 and 4, the word-initial contrast conditions are the more natural conditions, in which learning is hypothesized to be better, and the word-final contrast conditions are the less natural conditions, in which learning is hypothesized to be worse.

While Trained Contrast Position translates straightforwardly to expected naturalness across both experiments, Trained Neutralization Value does not translate straightforwardly to expected complexity, precisely because of the difference in non-critical consonants between the two experiments. Thus I do not wish to pair conditions that match in Trained Neutralization Value across the experiments. Instead, I want to pair the conditions that, within the natural conditions and within the unnatural conditions, match in expected complexity. Recall that in Experiment 3, with sonorant non-critical consonants, the simple conditions are those in which the Novel Nonconforming items feature voiceless stops (i.e. neutralizing-to-voiced conditions) and the complex conditions are those in which the Novel Conforming items feature voiced stops (i.e. neutralizing-to-voiceless conditions), while in Experiment 4, with voiceless fricative non-critical consonants, the reverse is true. Thus to pair conditions across the experiments, I match within conditions that shared Trained Contrast Position the neutralizing-to-voiced condition from Experiment 3 with the corresponding neutralizing-to-voiceless condition from Experiment 4. This pairs conditions with the same expected naturalness and the same expected complexity. Table 3.8 shows the four types of conditions that exist in both experiments and the specific conditions in each experiment that represent these types and are paired in the combined analysis.

Table 3.8: Pairing of Conditions Across Experiments 3 and 4

Condition Type	Experiment 3 Condition	Experiment 4 Condition
Word-Final Contrast Complex	**_#T…{T, D}# (*#D)_**	**#D…{T, D}# (*#T)**
Word-Final Contrast Simple	**#D…{T, D}# (*#T)**	**_#T…{T, D}# (*#D)_**
Word-Initial Contrast Complex	_#{T, D}…T# (*D#)_	#{T, D}…D# (*T#)
Word-Initial Contrast Simple	#{T, D}…D# (*T#)	_#{T, D}…T# (*D#)_

By combining the results of Experiments 3 and 4 as in Table 3.8, I can conduct the same analysis of Novel Nonconforming items to test the same hypotheses as in the individual analyses. In a regression model of Novel Nonconforming items, the effect of position-based substantive bias would again emerge as a significant main effect of Trained Contrast Position whereby acceptance rates of Novel Nonconforming items are higher in final contrast conditions than in initial contrast conditions. (Recall that in Experiment 3 the acceptance rate of Novel Nonconforming items was numerically higher in the final contrast condition than in the initial contrast condition within the neutralizing-to-voiced conditions but not within the neutralizing-to-voiceless conditions. In Experiment 4, the acceptance rates of Novel Nonconforming items were always numerically higher in the final contrast conditions than in their corresponding initial contrast conditions.) The effect of complexity bias would emerge as a significant main effect of the new factor Complexity (simple vs. complex) whereby acceptance rates of Novel Nonconforming items are higher in complex conditions than in simple conditions. This would mean that participants were more likely to erroneously accept words that did not belong to their language in conditions in which they had to master a more complex constraint than in conditions in which a simpler constraint was available to them. (In Experiments 3 and 4, the acceptance rates of Novel Nonconforming items were always numerically higher in the complex conditions than in their corresponding simple conditions.)

To test for the effects of substantive bias and complexity bias in both experiments together, I fit a logistic mixed-effects regression to the combined Novel Nonconforming items with response (accept or reject) as the dependent variable, Filler Consonant Type (sonorant vs. fricative), Trained Contrast Position (initial vs. final), Complexity (simple vs. complex), the interaction of Filler Consonant Type and Trained Contrast Position, and the interaction of Filler Consonant Type and Complexity as fixed effects, and random intercepts for subject and item.[35] The main effects of Trained Contrast Position and Complexity respectively test for the effects of position-based substantive bias and complexity bias, as mentioned above. I included the interactions of Filler Consonant Type with each of these two factors in case one or both types of bias had a stronger effect in one experiment than in the other. As the individual analyses showed,

---

[35] With full interaction structure, i.e. with all possible 2-way interactions and the 3-way interaction of all factors, the 3-way interaction of Filler Consonant Type, Trained Contrast Position, and Complexity, which would capture differences in the interaction of Trained Contrast Position and Complexity between the two experiments, is not significant ($p = 0.229$). Neither is the 2-way interaction of Trained Contrast Position and Complexity, whether the reference level of Filler Consonant Type is sonorant ($p = 0.179$) or fricative ($p = 0.718$). The 3-way interaction of Filler Consonant Type, Trained Contrast Position, and Complexity does not contribute significantly to the full model ($\chi^2(1) = 1.432$, $p = 0.231$). The 2-way interaction of Trained Contrast Position and Complexity does not contribute significantly to a model with all possible 2-way interactions of Filler Consonant Type, Trained Contrast Position, and Complexity ($\chi^2(1) = 0.486$, $p = 0.486$).

the results of Experiment 4 exhibited an effect of substantive bias while the results of Experiment 3 did not. Complexity bias emerged in both experiments, though not to the same degree: its effect was significant in Experiment 3 but only marginally significant in Experiment 4. Thus it was important to include the interactions of Filler Consonant Type with Trained Contrast Position and with Complexity.

The interaction of Filler Consonant Type and Trained Contrast Position was in the direction of a stronger effect of substantive bias in Experiment 4 (i.e. when filler consonants were fricatives), but it was not significant ($p = 0.229$). The interaction of Filler Consonant Type and Complexity was in the direction of a stronger effect of complexity bias in Experiment 3 (i.e. when filler consonants were sonorants), but it was not significant either ($p = 0.194$). Additionally, neither interaction contributed significantly to a model in which it was the only 2-way interaction, nor did the inclusion of both interactions significantly improve on the model with no interactions.[36] Consequently, I removed both of these interactions from the model, which meant that the main effects of Trained Contrast Position and Complexity represented effects across both experiments. The final model included the fixed factors Filler Consonant Type, Trained Contrast Position, and Complexity, with no interaction terms. The fixed effects of this model are given in Table 3.9.

There was no significant effect of Filler Consonant Type, meaning that there was no tendency for acceptance rates of Novel Nonconforming items to be overall higher in one experiment than in the other. There was a significant main effect of Trained Contrast Position

---

[36] The 2-way interaction of Filler Consonant Type and Trained Contrast Position does not contribute significantly to the model in which it is the only interaction ($\chi^2(1) = 1.389$, $p = 0.239$). The 2-way interaction of Filler Consonant Type and Complexity also does not contribute significantly to the model in which it is the only interaction ($\chi^2(1) = 1.626$, $p = 0.202$). The two 2-way interactions together do not significantly improve on the model with no interactions ($\chi^2(2) = 3.060$, $p = 0.217$).

such that, across both experiments, acceptance rates of Novel Nonconforming items were higher in final contrast conditions than in initial contrast conditions ($p = 0.030$). This means that, overall, participants trained on the obstruent voicing contrast word-finally were more likely to extend it to initial position than participants trained on the contrast word-initially were to extend it to final position, in line with the phonotactic implicational. When the results of the two voicing contrast experiments are combined, there is evidence for an effect of position-based substantive bias. Finally, there was a significant main effect of Complexity such that, across both experiments, acceptance rates of Novel Nonconforming items were higher in complex conditions than in simple conditions ($p < 0.001$). In other words, participants were overall better at rejecting nonconforming words when they could take advantage of a simpler constraint to master their language than when they had to learn a more complex constraint to master their language. Combining the results of the two voicing contrast experiments yields strong evidence for complexity bias.

Table 3.9: Combined Results of Experiments 3 and 4—Fixed Effects of the Novel Nonconforming Items Model

	β	p
Intercept	-1.155	<0.001***
Filler Consonant Type = fricative (vs. sonorant)	-0.058	0.784
Trained Contrast Position = final (vs. initial)	0.458	0.030*
Complexity = complex (vs. simple)	0.793	<0.001***

Since the main effect of Complexity has a larger coefficient and is more significant than the main effect of Trained Contrast Position, the two experiments together suggest that complexity bias is stronger than position-based substantive bias. The combined analysis largely confirms what I found broadly in the individual analyses of Experiments 3 and 4, namely, more compelling evidence for complexity bias than for substantive bias. In the individual analyses, the effect of complexity bias was always at least marginally significant, but the effect of position-based substantive bias was only significant in Experiment 4. The pooled analysis shows more directly that the effects of complexity bias are greater than the effects of substantive bias. This is consistent with Moreton & Pater (2012b)'s claim that complexity bias is robust while substantive bias is elusive and therefore likely subtle.

CHAPTER 4: EXPERIMENT 5: BIASED LEARNING OF WORD-FINAL OBSTRUENT VOICING ALTERNATIONS

This chapter reports on a collaboration with Jesse Zymet, Jacob Adams, Rachel Jacobson, Anthony Yates, Ann Zeng, and Robert Daland. This work was reported in part at the 2018 Annual Meeting of the Linguistic Society of America and the 36th West Coast Conference on Formal Linguistics.

Experiments 3 and 4 tested for substantive bias by comparing the learnability of phonotactic systems with different distributions of an obstruent voicing contrast. In Experiment 5, I also test for substantive bias using the phenomenon of obstruent voicing, but unlike Experiments 1–4, which were phonotactic learning experiments, this study is an alternation learning experiment. Additionally, in Experiment 5 the predictions of the substantive bias hypothesis are motivated by different factors.

Experiment 5 tests the relative learnability of word-final obstruent devoicing and word-final obstruent voicing. Word-final obstruent devoicing, commonly called final devoicing, is the phonological process whereby underlyingly voiced obstruents become voiceless in word-final position. It is cross-linguistically common, occurring in such languages as Dutch, German, Polish, Catalan, and Turkish (Lombardi 1991, Brockhaus 1995). Such languages feature an obstruent voicing contrast in their phonology, but word-finally the contrast is neutralized so that only voiceless obstruents occur. The German data in (1) exemplifies this pattern (data from Brockhaus 1995):

(1)	a.	*Bund*	[bʊnt]	‘league’	*Bunde*	[bʊnd-ə]	‘league-DAT.SG’
	b.	*bunt*	[bʊnt]	‘colorful’	*bunte*	[bʊnt-ə]	‘colorful-FEM.NOM.SG’

The minimal pair *Bunde* ‘league-DAT.SG’ vs. *bunte* ‘colorful-FEM.NOM.SG’ shows that voicing is contrastive in obstruents in German. When a voiced obstruent winds up in word-final position, however, as in *Bund* ‘league,’ it devoices. The allomorphs of the stem *Bund* in (1a) illustrate the alternations that underlyingly voiced obstruents exhibit, and the homophony of *Bund* ‘league’ and *bunt* ‘colorful’ shows that German’s obstruent voicing contrast is neutralized word-finally.

The complementary pattern, in which the voicing contrast is neutralized to *voiced* word-finally, is virtually unattested (Lombardi 1991, Blevins 2004, Gordon 2016). The only purported example of word-final voicing arises in Lezgian, a Nakh-Daghestanian language spoken in the Caucasus, and it is a marginal case, with final voicing applying only in monosyllabic nouns (Yu 2004).[37] The cross-linguistic prevalence of final devoicing and non-existence of final voicing is a well-known typological asymmetry. One possible explanation for this asymmetry is substantive bias. That is, learners acquiring language prefer final devoicing to final voicing because final devoicing is more natural than final voicing, and as a result languages frequently exhibit final devoicing but never exhibit final voicing.

In Experiments 3 and 4, I focused on positional naturalness in testing for substantive bias. I hypothesized that participants would be biased toward learning phonotactic systems with an obstruent voicing contrast only word-initially over phonotactic systems with an obstruent voicing contrast only word-finally. An obstruent voicing contrast word-finally should asymmetrically entail an obstruent voicing contrast word-initially since the voicing contrast is easier to perceive word-initially than word-finally and contrasts that exist in a given position should also exist in all

---

[37] Blevins (2004:109-110) dismisses a few other potential cases of final voicing.

positions in which they are easier to perceive. The reasoning behind the substantive bias hypothesis was rooted in perception, and more precisely in the relative perceptibility of obstruent voicing in different phonological contexts.

Though it was not the focus of my predictions, I did acknowledge a voicing-related substantive bias hypothesis. In addition to varying whether participants were exposed to an obstruent voicing contrast word-initially or word-finally, I varied whether the obstruents in the word-edge position without the contrast were voiced or voiceless. I considered that participants might perform better in conditions in which the training language "neutralized" to voiceless obstruents rather than voiced obstruents because voiced obstruents are more marked. I only touched on the phonetic grounding of this bias, however, and in any case the results of Experiments 3 and 4 taken together suggest that there was no effect of such a voicing-based substantive bias in those studies.

In the case of final devoicing vs. final voicing, however, the difference between the two phonological systems is only one of voicing value and not one of position. Both final devoicing and final voicing are natural with respect to the Steriadian implicationals about the distribution of contrasts across contexts. Both neutralize the obstruent voicing contrast in a position in which it is relatively difficult to perceive while the contrast is maintained in positions where it is easier to perceive. They differ in the direction of neutralization. The positional implicationals do not bear on this question; they only govern the presence or absence of the contrast.

Steriade (1997) does predict that, in the case of word-final neutralization of an obstruent voicing contrast, neutralization will always be to voiceless, but the reason lies in articulation rather than in perception. In word-final position, a voiced stop is more difficult to produce than a voiceless stop (Westbury & Keating 1986). Producing voicing requires a pressure differential across the glottis, with higher subglottal pressure and lower supraglottal pressure. This differential is harder to maintain when there is a complete closure in the vocal tract above the glottis. Voiceless stops do not require maintaining this pressure differential, so they are easier to produce (at least in word-edge position; see Westbury & Keating 1986). Consequently, final devoicing promotes articulatory ease by making all word-final stops of the type that is easier to produce in this position. Final voicing, on the other hand, would reduce articulatory ease by making all word-final stops of the type that is harder to produce in this position. Final devoicing is therefore phonetically natural while final voicing is not.

Experiment 5 tests for an articulatorily-rooted substantive bias effect by investigating how well final devoicing and final voicing alternations are learned in an artificial grammar learning paradigm. If learners are biased toward phonetically natural patterns—in this case, phonetically natural alternations—natural final devoicing should be learned better than unnatural final voicing. We also included a third training pattern that exhibited an exchange rule: voiced obstruents devoiced word-finally while voiceless obstruents voiced word-finally. This more complex pattern involving final devoicing and final voicing should be learned worse than either final devoicing or final voicing according to the complexity bias hypothesis.

To my knowledge, only two previous studies have tested final voicing alternations in an AGL paradigm. Myers & Padgett's (2014) Experiment 2 taught participants a final devoicing pattern using a singular-plural alternation (e.g. singular [ɡit]-plural [ɡid-i]). Like their Experiment 1, which was similar to Experiment 3, their Experiment 2 was designed to test domain generalization, that is, whether participants would extend phrase-final devoicing to word-final phrase-medial position. In the test phase, the task was to say whether a given plural-singular pair (e.g. [ɡid-i]-[ɡit]), with each form presented in the same frame sentence, belonged to the training language or not. Having been trained on the final devoicing pattern in phrase-final position, participants were more likely to accept phrase-final plural-singular pairs that conformed to the pattern than pairs that did not, demonstrating that they had learned final devoicing. Additionally, participants extended the final devoicing pattern from phrase-final position to word-final position, though here their responses did not conform to the pattern as often as they did for phrase-final position. Myers & Padgett's study showed that participants can learn final devoicing (and generalize it across domains).

In a study with Norwegian speakers, Lysvik (2018) tested the learnability of final devoicing and final voicing. He also taught the patterns using a singular-plural alternation (e.g. singular [musab]-plural [musap-u] in the final voicing condition). Participants first completed a forced-choice task in which they had to choose between two versions of a singular form (e.g. [rusid] or [rusit], where [rusid] is the correct choice in the final voicing condition). Then they completed a production task in which, given a plural form, they had to produce the corresponding singular form (e.g. given [musap-u], the correct singular form is [musab] in the final voicing condition).[38] In both tasks, participants demonstrated equal learning of final devoicing and final voicing. That is, there appeared to be no naturalness advantage for final devoicing. Lysvik's design differs from ours in several respects. Notably, in the training phase of his study, 75% of stems alternated while in our study, as detailed below, only one third of stems alternated in the training phase of the final voicing and final devoicing conditions. Additionally,

---

[38] In both the forced-choice task and the production task, participants also had to choose or produce plural forms. In these trials, there was no correct answer, since in the plural forms the critical stem-final consonant was not in word-final position. These trials constituted only 15% of trials in both test phases. For these results, which do not bear on the learnability of final voicing and devoicing, see Lysvik 2018.

he used a production task while we did not. Given the differences between our experiments, there is the potential for our results to differ.

With regard to the exchange pattern, the typology and the literature suggest that this alternation should be difficult to learn. Anderson & Browne (1973) note that such “exchange rules” are quite rare and thus probably hard to acquire. After reviewing reported cases, they conclude that purely phonological exchange rules (i.e. rules that refer simply to a phonological structural description) do not exist. Instead, exchange rules always refer to morphology or particular lexical classes. Thus there is no circular phonology. Moreton (1999) agrees and demonstrates that Classical Optimality Theory cannot compute circular phonological patterns such as exchanges or chain shifts. Intuitively, this works as follows: if the winning candidate for a given input is not the faithful candidate, the faithful candidate must be more marked than the winning candidate. For example, schematically, in our exchange pattern, T# is the winner for an input /D#/, so D# must be more marked than T#. However, it is also the case that D# is the winner for an input /T#/, which means that T# must be more marked than D#. This gives rise to a contradiction, and Classical OT cannot generate such a pattern.

Further evidence for the difficulty of circular patterns comes from studies showing that such alternations are not productive. For instance, the famous Taiwanese tone circle, whereby Tone 51 → Tone 55, 55 → 33, 33 → 21, and 21 → 51 in non-XP-final position, is not reliably applied to novel words by Taiwanese speakers (Zhang, Lai, & Turnbull-Sailor 2006, Zhang, Lai, & Sailor 2011). As Moreton (1999) points out, the Taiwanese tone circle does not disprove the non-existence of circular phonology since its environment makes reference to extraphonological (or at least phrase-level prosodic) properties. Even this not purely phonological process is imperfectly internalized by speakers, though.

Becker, Nevins, & Levine (2012), in their Experiments 4a and 4b, taught an artificial process of plural formation that could have been interpreted as an exchange rule. The authors describe the pattern as an umlaut process in which the plural is formed by reversing the value of the backness feature of the stressed vowel in the singular (e.g. singular [ˈzuməp] to plural [ˈziməp]). This appears to be a circular pattern: in the plural, stressed front vowels become back and stressed back vowels become front. However, in the artificial language, the only stressed stem vowels in singular forms were [u] and [e]; these respectively became [i] and [o] in the plural. Participants never encountered a true segmental exchange rule (e.g. [u] → [i] *and* [i] → [u]). The data they were exposed to are consistent with a featural exchange rule (i.e. [α back] → [−α back]) or with a plural formation pattern in which high vowels front ([u] → [i]) and mid vowels back ([e] → [o]). This second interpretation of the pattern is complex and might therefore be expected to be difficult to learn. In Becker, Nevins, & Levine's experiments, participants' rate of application of the umlaut process they had been trained on was always less than 50% (the task was written production). One possible explanation for this relatively poor performance is the aforementioned complexity: participants may have been trying to learn separate rules for high and mid vowels. Another possible explanation, more relevant to our study, is that exchange rules are hard to learn, even when they constitute a morphological process like plural formation and are thus not purely phonological.

The exchange pattern in Experiment 5 is ambiguous between a purely phonological process and a process triggered by a specific morpheme since, as will be explained in the following section, it was taught exclusively in the context of a plural-singular alternation. From the training data alone, it cannot be determined whether voiced obstruents devoice and voiceless obstruents voice word-finally in general or whether the exchange pattern only arises in nominal number marking. In designing this condition, we intended the exchange pattern to be true circular phonology, though we could not ensure that participants approached it as such. Given that the typology and experimental evidence suggest that even exchange patterns that are not purely phonological are hard to learn, we expected our exchange pattern to prove difficult whether participants treated it as pure phonology or not.

4.1 Method

4.1.1 Conditions

In Experiment 5, participants were trained on a phonological alternation related to the voicing of word-final obstruents and then tested on their ability to correctly apply that alternation. The alternation was taught as a plural-singular alternation. The plural form in the artificial language consisted of a stem followed by the suffix [-i] (cf. White 2013, Myers & Padgett 2014). The singular form consisted of just the stem, putting the stem-final consonant in word-final position. The plural form revealed the underlying voicing of stem-final consonants. In the singular form, the voicing of stem-final consonants sometimes differed from their underlying voicing, depending on the condition.

There were three conditions, which differed in the voicing behavior of obstruents word-finally. The alternation patterns of each condition are summarized and exemplified in Table 4.1. As the table shows, all three conditions featured three stem types: voiceless obstruent-final stems, voiced obstruent-final stems, and sonorant-final stems. In the Devoicing condition, obstruents neutralized to voiceless word-finally. Thus voiceless obstruents did not alternate, as demonstrated by the plural and singular forms of the voiceless-final stem *kinɛ́p*, but voiced obstruents devoiced in word-final position, as demonstrated by the plural and singular forms of the voiced-final stem *rɛkɛ́b*. In the Voicing condition, obstruents neutralized to voiced word-finally, so voiceless obstruents voiced in word-final position (e.g. *kinɛ́p-i ~ kinɛ́b*) while voiced obstruents did not alternate (e.g. *rɛkɛ́b-i ~ rɛkɛ́b*). The Devoicing and Voicing conditions differed minimally in the voicing value to which obstruents neutralized word-finally. Finally, in the Exchange condition, the voicing contrast was not neutralized word-finally; instead, both voiced and voiceless obstruents alternated. Voiceless obstruents voiced word-finally (e.g. *kinɛ́p-i ~ kinɛ́b*) while voiced obstruents devoiced word-finally (e.g. *rɛkɛ́b-i ~ rɛkɛ́p*). In all three conditions, there were also sonorant-final stems like *gipál* whose final consonants never alternated between the plural and the singular.

Table 4.1: Experiment 5 Training Conditions with Sample Training Items by Stem Type

Condition	**Voiceless-Final Stem**	**Voiced-Final Stem**	**Sonorant-Final Stem**
Devoicing	kinɛ́p-i ~ kinɛ́p	rɛkɛ́**b**-i ~ rɛkɛ́**p**	gipál-i ~ gipál
Voicing	kinɛ́**p**-i ~ kinɛ́**b**	rɛkɛ́b-i ~ rɛkɛ́b	gipál-i ~ gipál
Exchange	kinɛ́**p**-i ~ kinɛ́**b**	rɛkɛ́**b**-i ~ rɛkɛ́**p**	gipál-i ~ gipál

4.1.2 Materials

The training and test stems were nonce words of the form CVCV́C. Plural forms added the suffix [-i] without changing the original location of stress, giving CVCV́C-*i*. The stem vowels were drawn from [i ɛ ɑ oʊ u], but [i] never appeared as the second stem vowel, to avoid having two [i]s in a row in plural forms. The stem consonants were drawn from the sonorants [m n l ɹ] and the obstruents [p t k b d ɡ]. Within each stem type (voiceless-final, voiced-final, and sonorant-final), the two non-final consonants could be sonorants or obstruents. Within voiceless-final and voiced-final stems in both training and test, labial, coronal, and velar stops were equally represented as stem-final consonants. Within sonorant-final stems, the sonorants [m n l ɹ] were approximately equally represented as stem-final consonants.

In the training phase, participants were exposed to plural-singular pairs for 12 voiceless-final stems, 12 voiced-final stems, and 12 sonorant-final stems. The stems were the same across all three conditions, but the singular forms of the stems differed according to the voicing alternation active in each condition. For instance, the plural-singular pair for the stem *kinέp* was *kinέp-i* ~ *kinέp* in the Devoicing condition but *kinέp-i* ~ *kinέb* in the Voicing and Exchange conditions. In the test phase, participants were presented with a plural form and had to choose the correct singular form for their training language from two options, one alternating and one non-alternating. For obstruent-final stems, the two options differed in the voicing of the final stop. For example, for the plural form *dirúg-i*, the two singular options were non-alternating *dirúg* (correct in the Voicing condition) and alternating *dirúk* (correct in the Devoicing and Exchange conditions). For sonorant-final stems, which did not alternate in any of the three training languages, the alternating option was created by having [m]~[n] and [l]~[ɹ] alternate. Thus for the plural form *lipám-i*, the two singular options were non-alternating *lipám* (correct) and alternating *lipán* (incorrect). The test stems were the same in all three conditions and consisted of 12 voiceless-final stems, 12 voiced-final stems, and 12 sonorant-final stems. The choices for the singular form were also the same in all conditions, though which singular form was correct varied by condition. Within each stem type, half of the test stems were familiar, having appeared in the training phase, while half were novel. The full set of training and test stimuli for Experiment 5 are given in the Appendix.

The training and test items were recorded by a phonetically-trained female native speaker of American English who was naïve to the purpose of the experiment. All word-final stops were released, and word-final voiced stops were largely voiced during closure. Unstressed vowels were not reduced, and word-medial /t/s and /d/s that would normally tap in American English were not tapped. Otherwise, pronunciation was as in American English. The stimuli were recorded in a soundproof booth with a sampling rate of 22,050 Hz.

Following Kapatsinski 2013, the visual stimuli for the training and test phases were alien pictures taken from the website Sporepedia (www.spore.com/sporepedia). Individual pictures were multiplied and resized to make plural pictures.

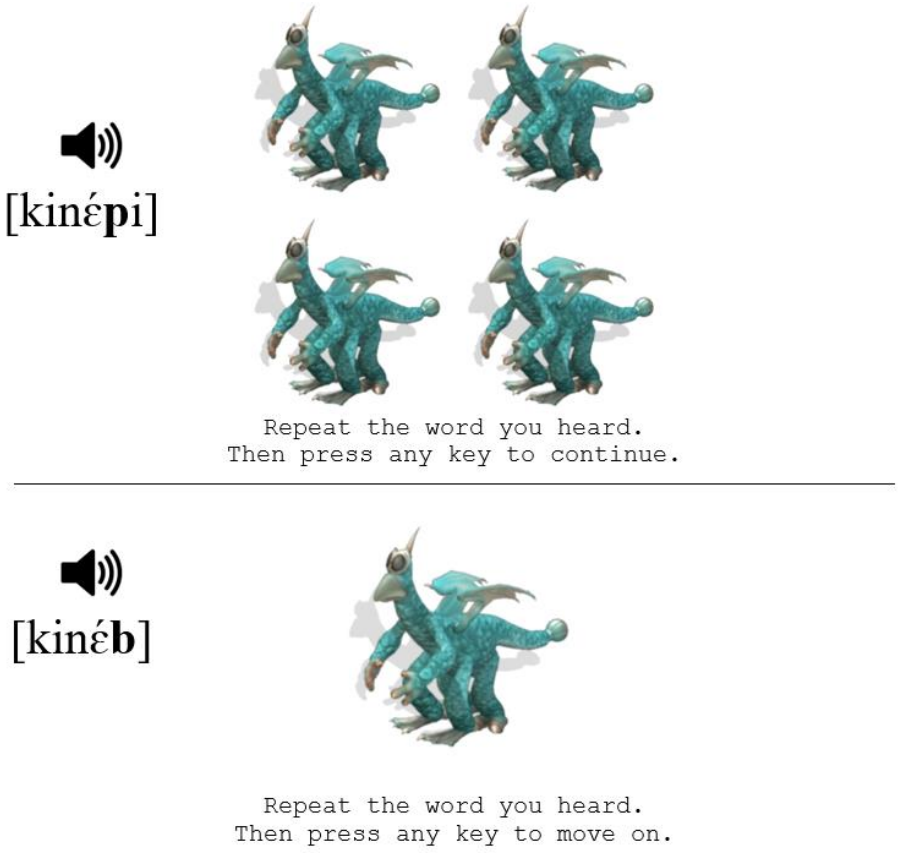
4.1.3 Procedure

Experiment 5 was implemented with the software E-Prime 2.0 and conducted in person. Before the experiment began, the experimenter asked the participant about their language background, including native language(s), languages spoken, and languages studied. The experiment was carried out in a soundproof booth. The participant was seated at a computer and wore headphones to hear the auditory stimuli. Labels with the numbers 1 and 2 were placed over the F and J keys, respectively, of the computer keyboard. In the test phase, the participant used these labeled keys to choose a singular form option, with 1 corresponding to the first option presented and 2 corresponding to the second. The experimenter launched the experiment, selected one of the three conditions (cycling through conditions with each successive participant), and stayed in the booth while the participant read the instructions for the training phase. The instructions told participants that they would be learning words from a new language. They would hear pairs of words and see alien pictures. The first word was how to say multiple aliens (i.e. the plural) of a particular species, and the second word was how to say one alien (i.e. the singular) of that species. They were told that in the second part of the experiment, they would be tested on what they thought the aliens were called. The instructions also asked participants to repeat each word after they heard it to help them learn. The experimenter asked whether the participant had any questions and then left the booth. To begin the experiment, participants pressed the space bar.

The training phase consisted of a single block of 36 plural-singular pairs. The training stems were randomly assigned to particular aliens each time the experiment was run, so the correspondence between stems and pictures was different for each participant. This was to avoid having a fixed correspondence between stems and aliens in which stems of a particular type were assigned to especially salient aliens, potentially making stems of that type especially noticeable and therefore easier to master. Each training trial began with a blank screen for 1000 ms. Then the plural picture, containing three, four, or six identical aliens, appeared. After 1000 ms, the plural form corresponding to the picture was played. Immediately after the auditory stimulus was presented, text appeared telling participants to repeat the word they had heard and then to press any key to continue. When participants pressed to continue, there was a blank screen for 1000 ms, and then the singular picture, featuring one alien of the same kind seen in the plural picture, appeared. After 1000 ms, the singular form corresponding to the picture was played. Immediately after the auditory stimulus was presented, text appeared telling participants to repeat the word they had heard and then to press any key to move on. By pressing any key, participants proceeded to the next training trial. Figure 4.1 schematizes one training trial.

After the training phrase, participants were allowed to take a break if they liked and were told to press the space bar to continue the experiment. Written instructions for the test phase then appeared, telling participants they would be tested on what they thought the aliens were called. They would first see a picture of multiple aliens (i.e. the plural) and hear what they were called. Then they would see a picture of a single alien (i.e. the singular) of the same kind and hear two different options for what to call the alien by itself. Participants were asked to press 1 if the first option sounded like the correct form and 2 if the second form sounded like the correct form. If they were not sure, participants were to make their best guess. Participants pressed the space bar to begin the test phase.

Figure 4.1: Event Sequence for One Training Trial

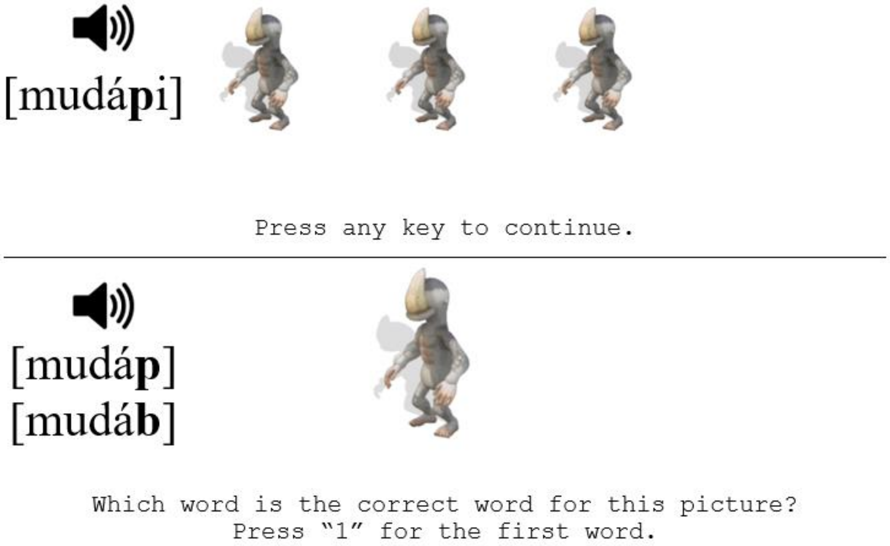


Repeat the word you heard.
Then press any key to continue.

Repeat the word you heard.
Then press any key to move on.

The test phase consisted of a single block of 36 trials. Half the trials featured familiar stems that the participants had been trained on while half featured novel stems they had not heard before. The familiar test stems were associated with the same alien pictures as in the training phase. The association between novel test stems and novel alien pictures was randomly generated for each participant so that there was no fixed correspondence of stems and pictures across participants. Each test trial began with a blank screen for 1000 ms. Then a plural picture, containing three, four, or six aliens, appeared. After 1000 ms, the plural form corresponding to the picture was played. After presentation of the auditory stimulus, text appeared telling participants to press any key to continue. There was again a blank screen for 1000 ms, and then the singular picture appeared along with the text: “Which word is the correct word for this picture? Press ‘1’ for the first word. Press ‘2’ for the second word.” After 1000 ms, the first option for the singular form played, and 500 ms later, the second option played. The order of the options was counterbalanced so that half the time the correct option for the given condition was presented first and half the time the incorrect option was presented first. When participants pressed 1 or 2 to make their choice, the experiment proceeded to the next test trial. Figure 4.2 schematizes one test trial. At the end of the test phase, text appeared thanking participants for participating.

Figure 4.2: Event Sequence for One Test Trial



4.1.4 Participants

The participants were native English speakers recruited through the UCLA Psychology Subject Pool. A total of 102 participants completed the experiment, and 27 were excluded post-hoc for reporting a native language other than English. This left 75 usable participants divided almost equally across the three conditions. No participant had more than negligible proficiency in any language with active word-final obstruent devoicing.

4.2 Predictions

As in Experiments 1, 3, and 4, it makes sense to predict that test phase performance should be better on familiar stems than on novel stems. For familiar test stems, participants were exposed to the correct singular form in training while for novel test stems they were not. Since memory can help for the familiar stems but not for the novel stems, participants should be worse at choosing the correct singular form for novel stems. That said, this familiarity effect does not always emerge in AGL experiments involving the learning of alternations (e.g. Finley 2012).

We initially made two bias-related predictions about the results of Experiment 5. The first prediction follows from the substantive bias hypothesis, which proposes that learners are biased toward phonetically natural phonological patterns. Since final devoicing is more natural than final voicing, performance should be better in the Devoicing condition than in the Voicing condition. If substantive bias is not active in the learning of final voicing alternations, performance should not differ between the Devoicing and Voicing conditions because the two alternations are simply mirror images of each other. The second prediction follows from the complexity bias hypothesis, which proposes that learners are biased against more complex phonological patterns. The Exchange condition is more complex than both the Devoicing and Voicing conditions because it involves both final devoicing and final voicing and requires keeping track of a stem-final consonant’s voicing in the plural to know what its voicing should be in the singular. Moreover, the pattern in the Exchange condition is a circular phonological process, a type of pattern that Optimality Theory cannot capture and that is believed not to exist (Moreton 1999). Consequently, we predicted that performance in the Exchange condition would be worse than in both the Devoicing and the Voicing conditions (see Pycha et al. 2003 for a study with three loosely analogous conditions involving vowel (dis)harmony and corresponding predictions).

There is another possible prediction from complexity bias that concerns the relative learnability of the Devoicing and Voicing patterns. Recall that in Experiment 3, a phonotactic learning experiment, one way in which the training languages differed was in whether the word-edge position that did not allow an obstruent voicing contrast featured voiced or voiceless obstruents. Consider just the two conditions in which obstruents were neutralized word-finally. In one condition, word-final obstruents were always voiceless (*D#) while in the other, word-final obstruents were always voiced (*T#). Participants were better at rejecting words that did not belong in their training language in the neutralizing-to-voiced (*T#) language than in the neutralizing-to-voiceless (*D#) language even though the greater markedness of voiced obstruents might have led us to expect that voiced obstruent-final words would be easier to reject. I proposed that this result was due to complexity bias and depended on the nature of the filler consonants in the training and test items. In both conditions, voiced sonorants were permitted word-finally. Thus to exclude final voiceless obstruents in the neutralizing-to-voiced language, participants had the option of just learning the one-feature constraint *[–voice]#. To learn to exclude final voiced obstruents, but not final voiced sonorants, in the neutralizing-to-voiceless language, participants necessarily had to learn the two-feature constraint *[–son, +voice]#. Since the phonotactic constraint needed to master the neutralizing-to-voiceless language was formally more complex than the constraint needed to master the neutralizing-to-voiced language, the neutralizing-to-voiced language was learned better, and this difference in learnability supports complexity bias. Experiment 4 provided further evidence for this account by demonstrating that changing the voicing of the filler consonants and therefore switching which training languages required a more complex constraint to learn caused the relative learnability of corresponding neutralizing-to-voiceless and neutralizing-to-voiced languages to switch.

From the point of view of a constraint-based grammar, Experiment 5 is analogous to the word-final neutralization conditions of Experiment 3. Experiment 5 involves alternation learning and Experiment 3 phonotactic learning, but in both cases participants must learn constraints on the consonants that are permitted in word-final position. In Experiment 5, this enables them to know how to correctly form the singular form of a given stem while in Experiment 3 it enables them to know which words could belong to their language and which words could not. Experiment 5 is also analogous to Experiment 3 in that the training and test phases included fillers that showed that sonorants could occur in word-final position (e.g. the plural-singular pair *gipál-i* ~ *gipál*). As a result, the constraints that participants had to learn in the Voicing and Devoicing conditions of Experiment 5 were in fact identical to the constraints that participants had to learn in the neutralizing-to-voiced (*T#) and neutralizing-to-voiceless (*D#) conditions, respectively. In the Voicing condition, participants needed to learn a constraint to drive final voicing, exemplified by a plural-singular pair like *kinɛ́p-i* ~ *kinɛ́b*. Since voiced obstruents and (voiced) sonorants were both allowed word-finally, participants could learn the one-feature constraint *[–voice]# to know that voiceless obstruents were disallowed word-finally and voiceless-final stems consequently had to voice in the singular. In the Devoicing condition, participants needed to learn a constraint to drive final devoicing, exemplified by a plural-singular pair like *rɛkɛ́b-i ~ rɛkɛ́p*. Voiceless obstruents were allowed word-finally, but so were (voiced) sonorants, so participants had to learn the two-feature constraint *[–son, +voice]# to know that voiced obstruents were disallowed word-finally and voiced-final stems had to devoice in the singular. They could not just learn *[+voice]# because this constraint would ban word-final sonorants, which were in fact permitted.[39]

Thus, as in Experiment 3, the presence of word-final sonorants makes the constraint needed to learn one pattern simpler than the constraint needed to learn the other pattern. In the Voicing condition, participants can learn a simpler constraint, namely *[–voice]#, to master their pattern while in the Devoicing condition participants must learn a more complex constraint, namely *[–son, +voice]#, to master their pattern. This difference in the complexity of the constraint needed to drive the alternation predicts that final voicing should be learned better than final devoicing. That is, performance should be better in the Voicing condition than in the Devoicing condition. This is precisely the opposite of the prediction from substantive bias, laid out above, whereby performance is expected to be better in the Devoicing condition than in the Voicing condition.

---

[39] Experiment 5 does differ from Experiment 3 because it is an alternation learning experiment and because its test phase involves a choice between two proposed singular forms. In Experiment 3, if participants learned the constraint *[+voice]# in the final devoicing analogue condition (*#{T, D}…T#*), then in the test phase they could (mistakenly) reject sonorant-final words as not belonging to their language. In Experiment 5, on the other hand, if participants learn *[+voice]# in the Devoicing condition, they must still choose a singular form for sonorant-final stems in the test phase. Native English-speaking participants presumably have an inviolable constraint *[+son, –voice] and thus will not entertain the possibility of devoicing a stem-final sonorant in word-final position. More importantly, there is no such option: both singular form choices for sonorant-final stems end in voiced sonorants. If participants in the Devoicing condition seek to apply *[+voice]# in the test phase, they will perform perfectly on obstruent-final stems. For the sonorant-final stems, they could either guess, yielding chance performance, or, more probably, choose the non-alternating singular option since *[+voice]# cannot be satisfied. By this reasoning, learning *[+voice]# in Experiment 5 could still lead to quite good performance in the test phase while learning *[+voice]# in Experiment 3 would have harmed performance in the test phase. All this said, in both experiments participants should be learning the constraints on the artificial language they are learning in the training phase, before they know what choices they will have in the test phase. From this perspective, participants in the relevant conditions in Experiments 3 and 5 are faced with the same distribution of voiced stops, voiceless stops, and voiced sonorants, and the constraint *[+voice]# is not consistent with the data they have encountered.

4.3 Results

Figure 4.3 shows the rates of correct responses on familiar and novel test stems of each type in each condition. In each of the three conditions, there are sonorant-final, voiced-final, and voiceless-final stems, as evidenced by the plural forms. The stem type(s) that alternate between the plural and the singular in each condition are outlined in black. For instance, in the Devoicing condition, voiced-final stems alternate, since they devoice word-finally in the singular (e.g. *rɛkɛ́b-i* ~ *rɛkɛ́p*). Within each stem type in a given condition, half the test stems are familiar and half are novel. The correct response rate for the familiar stems is represented by the dark gray bar on the left, and the correct response rate for the novel stems is represented by the light gray bar on the right.

Figure 4.3: % Correct Responses on Test Stems by Condition, Stem Type, and Familiarity

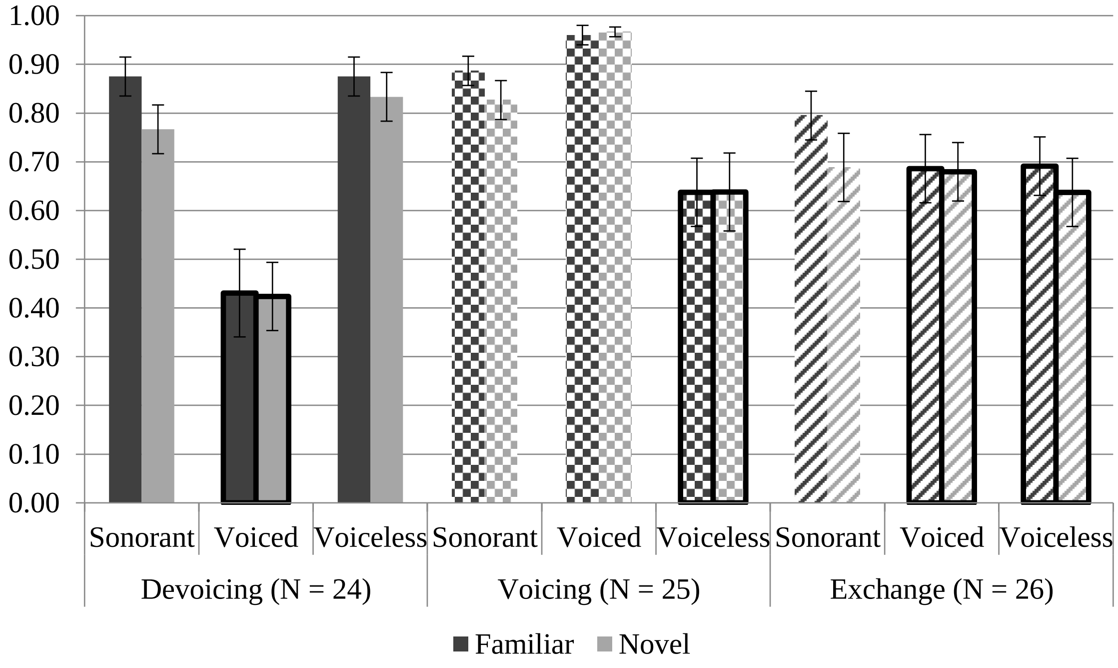


Figure 4.3 shows that the correct response rates for novel stems were for the most part quite similar to the correct response rates for familiar stems. Where there is a difference, the

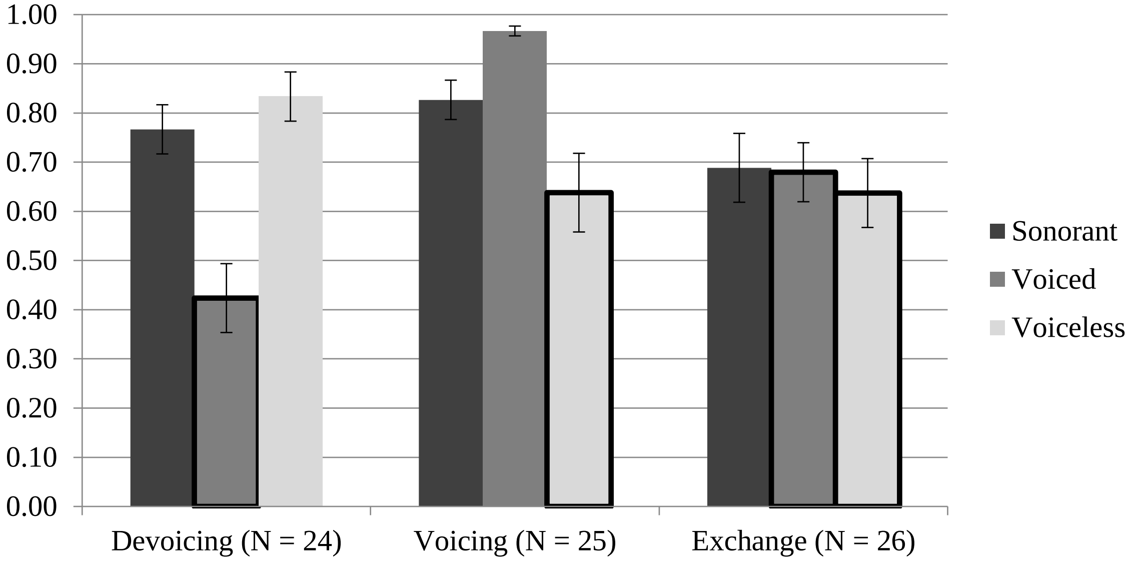
correct response rate is generally higher for familiar stems than for novel stems, as expected, but most of these differences do not look significant. I fit a mixed-effects logistic regression to the test items with response (correct or incorrect) as the dependent variable, Familiarity (familiar vs. novel) as a fixed effect, random intercepts for subject and item, and by-subject random slopes for Familiarity. The direction of the main effect of Familiarity was such that responses for novel stems were less likely to be correct than responses for familiar stems, but the effect was not significant ($\beta = -0.253$; $p = 0.108$). Overall, then, there was a tendency for participants to be better at choosing the correct singular form for stems they had been exposed to in the training phase, but this effect was not strong enough to reach significance. That is, participants were practically as good at choosing the correct singular form for novel stems as they were for familiar stems, with exactly how good they were depending on condition and stem type.

It is the novel stems that test how well participants have learned the alternation pattern they were trained on, since for these stems participants cannot rely on memory but must generalize to items they have not encountered before. Thus to compare performance in the different conditions, I analyze only the responses for novel stems. Figure 4.4 shows correct response rates for novel test stems by condition and stem type. Bars representing the stem type(s) that alternate in a given condition are outlined in black.

I first tested whether participants in each condition could apply the alternation they had been trained on at above chance levels. I fit a mixed-effects logistic regression to the novel items with response (correct or incorrect) as the dependent variable, Condition (Devoicing vs. Voicing vs. Exchange), Alternating (alternating vs. non-alternating), and their interaction as fixed effects, random intercepts for subjects and items, and by-subject random slopes for Alternating. The reference level of Alternating was always alternating, and I changed the reference level of

Condition so that in each model the intercept represented the correct response rate on alternating stems in one of the three conditions.

Figure 4.4: % Correct on Novel Stems by Condition and Stem Type



When the reference condition was Devoicing, the intercept represented the correct response rate on voiced-final stems in this condition. Participants had been trained to apply final devoicing to these stems. The intercept was negative but did not differ significantly from chance (β = -0.784; *p* = 0.174). This means that participants in the Devoicing condition did not learn to devoice word-final voiced obstruents.

When the reference condition was Voicing, the intercept represented the correct response rate on voiceless-final stems in this condition. Participants had been trained to apply final voicing to these stems. The intercept was significantly above chance (β = 1.216; *p* = 0.042). This means that participants in the Voicing condition did learn to voice word-final voiceless obstruents.

Finally, when the reference condition was Exchange, the intercept represented the correct response rate on all alternating stems, voiceless-final and voiced-final, in this condition.

Participants had been trained to apply final devoicing to voiced-final stems and final voicing to voiceless-final stems. The intercept was marginally significantly above chance ($\beta = 0.967$; $p = 0.066$). This means that participants in the Exchange condition weakly learned to apply the exchange rule (i.e. devoicing of word-final voiced stops and voicing of word-final voiceless stops).

To sum up these results, participants failed to learn natural final devoicing, as evidenced by their inability to apply this alternation at better than chance. On the other hand, participants did learn unnatural final voicing, and they very nearly learned the circular alternation pattern that combined final devoicing and voicing.

Table 4.2: Experiment 5—Fixed Effects of the Novel Items Model

	β	p
Intercept	-0.784	0.174
Condition = Voicing (vs. Devoicing)	1.999	0.015*
Condition = Exchange (vs. Devoicing)	1.751	0.020*
Alternating = non-alternating (vs. alternating)	2.955	<0.001***
Condition = Voicing × Alternating = non-alternating	-1.221	0.259
Condition = Exchange × Alternating = non-alternating	-2.355	0.019*

To compare performance between conditions directly, we can consider the regression of novel items in which the reference level for Condition was Devoicing. Table 4.2 gives the full set of fixed effects for this model. There was a significant main effect of Condition such that both performance in the Voicing condition ($\beta = 1.999$; $p = 0.015$) and performance in the Exchange condition ($\beta = 1.751$; $p = 0.020$) were better than performance in the Devoicing condition. There was also a significant main effect of Alternating whereby performance on non-alternating stems was better than performance on alternating stems ($\beta = 2.955$; $p < 0.001$). Finally, there was a significant interaction of Condition and Alternating whereby performance on non-alternating stems in the Exchange condition was not as much better than performance on alternating stems as otherwise predicted by the model ($\beta = -2.355$; $p = 0.019$). Importantly, there was no significant interaction of Condition and Alternating involving the Voicing condition. Thus the significant main effect of Condition can be interpreted as overall superior performance in the Voicing condition compared to the Devoicing condition. This can be seen in Figure 4.4, which shows that Voicing participants had higher correct response rates than Devoicing participants for both non-alternating and alternating stems.[40] Participants were also overall better at non-alternating stems, which is expected, except that participants in the Exchange condition did not perform as much better on non-alternating stems, relative to alternating stems. This can also be seen in Figure 4.4: the correct response rates for non-alternating stems are markedly higher than the correct response rates for alternating stems in Devoicing and Voicing, but not in Exchange, where the correct response rates for alternating and non-alternating stems are similar. The significant main effect of Condition whereby performance in Exchange is better than in Devoicing taken together with the significant interaction of Condition and Alternating involving the Exchange condition can be interpreted to mean that, on alternating stems, performance in Exchange was superior to that in Devoicing.[41]

---

[40] When the reference level of Alternating is non-alternating instead of alternating, the main effect of Condition such that performance in Voicing is better than in Devoicing is no longer significant ($\beta = 0.779$; $p = 0.174$). The lack of a main effect of Condition when the reference level of Alternating is non-alternating means that within non-alternating test stems, performance in the Voicing condition was not significantly better than in the Devoicing condition, though the difference in performance tends in that direction.

[41] This interpretation is confirmed by the fact that when the reference level of Alternating is changed to non-alternating, the main effect of Condition such that performance in Exchange is better than in Devoicing is no longer significant and is indeed in the opposite direction ($\beta = -0.604$; $p = 0.285$).

4.4 Discussion

The results of Experiment 5 are unexpected given our first two predictions. The substantive bias hypothesis predicted that final devoicing would be learned better than final voicing, but this was not the case. In the absence of substantive bias, we expected final devoicing to be learned no better than final voicing, or, put another way, for final voicing to be learned as well as final devoicing. This was not the case either. Instead, performance in the Voicing condition was better than performance in the Devoicing condition, meaning that phonetically unnatural final voicing was learned better than phonetically natural final devoicing. The complexity bias hypothesis predicted that the exchange pattern would be learned worse than both final devoicing and final voicing, but this was not true. Participants in the Exchange condition marginally learned to apply their alternation pattern, falling just short of participants in the Voicing condition and besting participants in the Devoicing condition, who did not learn to apply their alternation at all. Additionally, the regression analysis showed that Exchange participants outperformed Devoicing participants on alternating stems. Thus neither of our first two predictions were borne out.

Because final devoicing was not learned better than final voicing, the results do not provide any support for substantive bias. Although it is more natural, from an articulatory standpoint, to neutralize a voicing contrast to voiceless word-finally than to voiced, participants did not exhibit a preference for or a bias toward final devoicing over final voicing. Participants did not just learn both neutralization patterns to an equal degree, though. Not only was final devoicing not learned better than final voicing, but it was actually learned worse. The absence of a substantive bias effect does not predict the result whereby final voicing is learned *better* than final devoicing. However, the third hypothesis laid out in Section 4.2, which follows from complexity bias and the presence of sonorant-final stems in the stimuli, predicts exactly this result. Because all conditions featured stem-final voiced sonorants that did not alternate from the plural (word-medial position) to the singular (word-final position) (e.g. *gipál-i* ~ *gipál*), participants in the Devoicing condition had to learn the two-feature constraint *[–son, +voice]# to drive their word-final obstruent devoicing alternation while still admitting word-final voiced sonorants. Participants in the Voicing condition, on the other hand, could learn the one-feature constraint *[–voice]# to drive their word-final obstruent voicing alternation. By virtue of being formally simpler, the Voicing pattern should be learned better than the Devoicing pattern, and this is in fact the result we obtained. Complexity bias, in conjunction with the nature of the filler items, provides an account for an initially unexpected result of Experiment 5. This account is consistent with my proposed explanation for the results of Experiment 3, a phonotactic learning study that obtained a similar result. Experiment 5 thus joins Experiments 3 and 4 in yielding further support for complexity bias.

Recall that Lysvik (2018) found equal learning of final voicing and final devoicing in his AGL study with Norwegian speakers. This result also fails to provide support for the articulatorily-rooted substantive bias that Experiment 5 tested for. Unlike our stimuli, Lysvik's contained no filler items. All training and test stems ended in voiced or voiceless stops (Julian Lysvik, p.c.). Since there were no sonorant-final stems, there was no way in which the final voicing pattern was simpler than the final devoicing pattern and no way for the complexity bias found in our study to emerge. Thus his result is consistent with ours and is what we would predict in light of the effects uncovered in our experiment.

The second unexpected result of Experiment 5 was the surprisingly good performance in the Exchange condition. We expected the exchange pattern to be learned less well than both final devoicing and final voicing, but instead participants succeeded in applying the exchange pattern to novel stems at marginally above chance levels. They also learned their alternations better than Devoicing participants learned final devoicing. Complexity bias in phonological learning is well-supported, and other results in Experiments 3, 4, and 5 provide evidence for complexity bias. The exchange pattern in Experiment 5 cannot be mastered by learning a single constraint like *[–voice]# or *[–son, +voice]#. There is no generalization to be made about the types of consonants that can occur in word-final position, as there is in Devoicing and Voicing. Instead, the voicing of the stem-final consonant in the singular depends on its voicing in the plural. The learner must refer to the plural form in order to choose the correct singular form, which is not the case in Devoicing and Voicing.

A possible explanation for the unexpectedly high correct response rates in the Exchange condition is that Exchange participants heard twice as many alternating stems in the training phase. This was because in all three conditions there were three stem types (sonorant-final, voiced-final, and voiceless-final), and in the Exchange condition two out of three stem types alternated while in the Devoicing and Voicing conditions only one out of the three alternated. The experimental design ensured equal representation of stem types but, consequently, unequal proportions of alternating stems across conditions. In the Exchange condition, two thirds of plural-singular pairs in the training phase exhibited alternation, and this may have made participants more likely to choose an alternating option in the test phase.[42] In contrast, in the

---

[42] Indeed, Exchange participants chose the correct, i.e. the alternating, response on voiced-final and voiceless-final stems about two thirds of the time, as shown in Figure 4.4. However, it is not the case that hearing two out of three stems alternate in training made Exchange participants choose the alternating option for two out of three stems in test. They did so for stem types that were supposed to alternate (voiced-final and voiceless-final stems), but for sonorant-final stems they also chose the correct response around two thirds of the time, meaning that they chose the *non*-alternating response two thirds of the time. Exchange participants were able to treat non-alternating sonorant-final stems differently from alternating obstruent-final stems; they did not simply alternate two thirds of the time across stem types in test.

Devoicing and Voicing conditions, two thirds of plural-singular pairs in the training phase did *not* exhibit alternation, which may have made participants in these conditions less likely to choose an alternating option in test, all else being equal. From this perspective, the Exchange condition and the Devoicing and Voicing conditions are not comparable for the purposes of testing for an effect of complexity bias. The Exchange pattern was more complex, but in a way participants were also given more exposure to the pattern than they were in the Devoicing and Voicing conditions. Consequently, their high correct response rates do not necessarily constitute evidence against complexity bias.

Another possible explanation, not necessarily mutually exclusive from the first, is that participants in the Exchange condition engaged in fairly explicit learning, perhaps more explicit than in the Devoicing and Voicing conditions. While OT cannot capture the circular phonological alternation in the Exchange condition, the pattern is stated rather easily as a rule: if the stem-final consonant is an obstruent, switch its voicing from the plural to the singular. Participants may have been able to identify and apply such a rule in more linguistically naïve terms. They may even have taken a domain-general pattern-learning approach to the task, in which case their learning may not have been phonological. If so, their ability to learn the Exchange pattern does not mean circular phonology in natural languages is learnable.

In Experiment 5, a process of final devoicing was learned no better, and in fact was learned worse, than a process of final voicing, yielding no support for substantively biased phonological learning based on articulatory ease. The fact that final voicing was learned better than final devoicing instead provides support for complexity bias, since due to the sonorant-final filler stems, final voicing could be driven by a simpler constraint than final devoicing. This result is the alternation learning equivalent of the phonotactic learning result in Experiment 3 whereby a static “final voicing” language (#{T, D}…D#) was learned better than a static “final devoicing” language (*#{T, D}…T#*), when sonorants also occurred word-finally.

Experiment 3 and especially Experiment 4 also yielded some support for substantive bias insofar as participants exposed to an obstruent voicing contrast word-finally tended to extend that contrast to initial position more than participants exposed to the contrast word-initially extended it to final position. Experiment 5, on the other hand, yielded no evidence for substantive bias, but it did not test for the same substantive bias that emerged to a degree in the phonotactic learning experiments. For Experiments 3 and 4, I distinguished the position-based substantive bias hypothesis, which predicted asymmetric positional extension of the voicing contrast, and the voicing-based substantive bias hypothesis, which predicted more extension from voiced stops to voiceless stops than vice versa. The voicing-based substantive bias effect never emerged in Experiments 3 and 4, and it is this substantive bias hypothesis that Experiment 5 tests for. Thus the lack of an effect in the alternation learning experiment is consistent with the results of the phonotactic learning experiments. Learners in these studies do not seem to be biased by the relative markedness of voiced stops relative to voiceless stops. In Experiment 5, as in Experiment 3, the voicing-based substantive bias hypothesis and the complexity bias hypothesis make directly opposite predictions and so are effectively pitted against each other. Only one of the two bias effects can emerge. If both types of bias are active in learning, then presumably the one that is stronger is the one that will manifest an effect. It may be that substantive bias, as Moreton & Pater (2012b) have suggested, is simply weak in its effects and therefore the more robust complexity bias wins out over it.

The Exchange condition in Experient 5 was included to test a different sort of complexity bias than the one that Experiments 3 and 4 were able to test, and its pattern was learned surprisingly well. As discussed above, though, the Exchange condition differed from the Devoicing and Voicing conditions in more than just the nature of the alternation, and so performance in this condition cannot reliably bear on the complexity bias hypothesis.

Finally, while it was not the focus of Experiment 5, participants' better performance on non-alternating stems, at least in the Devoicing and Voicing conditions, is worth discussing in the context of previous studies. Setting aside the Exchange condition, in which participants were exposed to twice as many alternating stems in training, participants' superior ability to choose the correct singular form for stem types that did not alternate in their condition can be attributed to a bias against alternation. That is, within a condition, when participants had to apply an alternation in order to select the correct singular form, they performed worse despite having been trained on precisely this alternation. This anti-alternation bias is also a bias toward paradigm uniformity: if learners prefer that stems have the same form throughout a morphological paradigm, they should be worse at choosing the correct singular form when doing requires endorsing a non-uniform paradigm (e.g. *rεkέ**b**-i* ~ *rεkέ**p***), which was in fact the case in the Devoicing and Voicing conditions.

Previous work has also found a bias against alternation and a preference for paradigm uniformity, in various forms. Experiments have found a bias toward uniform paradigms in children. In an AGL study, English-learning 4-year-olds avoided applying repairs to difficult consonant clusters when doing so would create a non-uniform nominal paradigm (e.g. avoiding repairing the cluster [mf.d] in the plural form [næ**mf-d**əl] when the singular [næmf] is attested but simplifying monomorphemic [gʌ**mfd**εp] to [gʌmdεt]) (Tessier 2012). In a natural language study, Korean-learning 4- to 7-year-olds produced inflected verb forms with the incorrect stem allomorph or chose alternative inflections that allowed them to use a different stem allomorph, all in order to maintain more uniform verbal paradigms (Do 2018). Similarly, Dutch-learning 2- to 7-year-olds produced many incorrect plural noun forms by using the singular allomorph in the plural (e.g. *[bɛt-ən] instead of [bɛd-ən] for the plural of [bɛt] 'bed'; Dutch has final devoicing) (Kerkhoff 2004). Depending on age group, they failed to produce alternating plural forms between about 25% and 50% of the time, instead opting to keep the nominal paradigm uniform. Kerkhoff also wug-tested the Dutch children, asking them to produce the plural forms of non-words. Only a little over a quarter of her participants produced any alternating plural forms in the wug test, and these participants produced alternating plural forms only 10% of the time, less often than stems alternate in the Dutch lexicon. This further suggests that children have a preference for paradigm uniformity.

Children have also been observed to naturally produce forms that are phonotactically illegal in their native language but that preserve paradigm uniformity. For instance, a Greek-learning 4-year-old produced [ˈexete] 'you (pl) have,' with an illegal [xe] sequence, instead of the correct form [ˈeçete] because of the form [ˈexo] 'I have' (Kazazis 1969). In another case, between the ages of two and four an American English-learning child produced [sɪtin] *sitting* instead of [sɪɾɪŋ], with a tap, because of the base form [sɪt] *sit* (Bernhardt & Stemberger 1998). In typical American English adult speech a /t/ is produced as a tap when preceded by a stressed vowel and followed by an unstressed vowel; a [t] would not occur in this environment.

Other evidence for a preference for uniform paradigms comes from AGL experiments that teach unfamiliar alternations. In Wilson 2006, participants were taught a velar palatalization alternation but applied it at less than a 50% rate in the test phase. In Peperkamp, Skoruppa, & Dupoux 2006, French-speaking participants were taught intervocalic obstruent voicing alternations but applied them less than 50% of the time on familiar items and around 25% of the time on novel items. In Experiments 4a and 4b in Becker, Nevins, & Levine 2012, participants were taught an umlaut process to mark the plural (e.g. singular [ˈzuməp] to plural [ˈziməp]) but applied it less than 50% of the time in the test phase. Among other reasons, the low rates of application of a trained alternation in these studies may be due to the fact that they were production tasks (oral in Wilson 2006 and Peperkamp, Skoruppa, & Dupoux 2006 and written in Becker, Nevins, & Levine 2012) rather than forced choice tasks and that, in the case of Becker, Nevins, & Levine 2012, the alternation was stem-internal and not triggered by a phonological environment created through affixation or word concatenation. Even so, these results point to a baseline reluctance to alternate on the part of learners that is difficult to overcome even with explicit training.

As discussed in the introduction to this chapter, Myers & Padgett's (2014) Experiment 2 taught a final devoicing pattern in the context of a singular-plural alternation and tested generalization of the pattern from phrase-final to word-final position. Myers & Padgett explicitly predicted that, independent of their learning of the final devoicing pattern and their domain generalization, participants would prefer non-alternating singular-plural pairs (e.g. [tepi-i]-[tep]) over alternating singular-plural pairs (e.g. [teb-i]-[tep]). They predicted that this would be the case across-the-board, including for singular-plural pairs that did not conform to the trained final devoicing pattern and therefore should always have been rejected in the test phase (i.e. non-alternating pairs like [teb-i]-[teb] would be accepted more than alternating pairs like [tep-i]-[teb] even though both were nonconforming). Their prediction of a preference for non-alternation was borne out: participants accepted non-alternating singular-plural pairs more than non-alternating ones, though the difference was only significant in word-final (domain generalized) position.

This finding provides some support for a paradigm uniformity bias that is orthogonal to learning of the actual trained pattern.

Myers & Padgett do not examine the effect of (non-)alternating status within the test pairs that conformed to final devoicing (i.e. test pairs of the type participants had been trained on), which would be the more direct comparison to Experiment 5. However, within just conforming test pairs, the acceptance rates were always numerically higher for pairs in which the stem did not alternate (e.g. [tep-i]-[tep]) than for pairs in which it did (e.g. [teb-i]-[tep]), and this difference looks to be significant in word-final position, as it is across all test pairs (both conforming and non-conforming). This result would be parallel to the Experiment 5 result whereby participants in the Devoicing and Voicing conditions were better at choosing the correct singular form when the stem did not alternate than when it did. Additionally, participants in Myers & Padgett's experiment appeared to be better at discriminating conforming pairs from nonconforming pairs with the same underlying stem when the stem did not alternate in the conforming pair (e.g. [tep-i]-[tep] (belongs), [tep-i]-[teb] (does not belong)) than when the stem alternated in the conforming pair (e.g. [teb-i]-[tep] (belongs), [teb-i]-[teb] (does not belong)). That is, discrimination was better when the conforming pair exhibited paradigm uniformity than when the conforming pair did not. Myers & Padgett did not make this direct comparison either, though.

Summing up, Devoicing and Voicing participants' better performance on non-alternating stems in Experiment 5 provides additional experimental support for a bias against alternation and a preference for paradigm uniformity.

CHAPTER 5: CONCLUSION

5.1 Summary of Findings

This dissertation has reported on the results of five AGL experiments, four phonotactic learning experiments and one alternation learning experiment, all of which were designed to test for substantive bias. The phonotactic learning experiments used a novel design and tested whether participants would reproduce phonotactic implicationals found in the phonological typology and motivated by phonetic naturalness. The first set of two experiments tested the phonotactic implicational whereby major place contrasts in word-final position entail those same contrasts in word-initial position while major place contrasts in word-initial position do not entail those contrasts in word-final position. Experiment 1 exposed participants to a three-way labial/coronal/velar contrast in stops in one word-edge position and only coronal stops in the other position. Because in everyday speech major place contrasts are more perceptible in word-initial position than in word-final position, it is phonetically unnatural for a language to exhibit these contrasts word-finally but not word-initially. Consequently, I predicted greater extension of the place contrasts from final to initial position than from initial to final position. This prediction was not borne out, however: participants extended equally in both directions. Their behavior provided no support for substantive bias. In Experiment 2, I eliminated familiar (i.e. trained) items from the test phase to try to foster more implicit learning, but there was again no asymmetry in participants' willingness to extend the place contrasts to the other word-edge position. The phonotactic implicational about the distribution of major place contrasts was not reproduced experimentally, and this set of experiments found no evidence for substantive bias.

The second set of two experiments tested the phonotactic implicational whereby an obstruent voicing contrast in word-final position entails the same contrast in word-initial position but the reverse is not true. As with the place contrasts, the voicing contrast in obstruents (in particular, in stops) is more perceptible word-initially than word-finally, so the prediction from substantive bias was that participants would extend a stop voicing contrast more from word-final to word-initial position than from word-initial to word-final position. Experiments 3 and 4 both exposed participants to a stop voicing contrast in one word-edge position and to only voiced or voiceless stops in the other position. Conditions thus varied in the position exhibiting the contrast and in the neutralization value of [voice] in the position without the contrast. In Experiment 3, participants extended the voicing contrast more from word-final to word-initial position in the neutralizing-to-voiced conditions but not in the neutralizing-to-voiceless conditions. As a result, the experiment did not yield convincing support for substantive bias; no effect of substantive bias emerged in the statistical model of the results. In Experiment 4, participants did consistently extend the voicing contrast more from word-final to word-initial position, and the effect of substantive bias was significant. Together, the two experiments represent unreliable reproduction of the phonotactic implicational about the distribution of the stop voicing contrast and mixed evidence for the position-based substantive bias that could give rise to behavior conforming to the implicational.

Experiments 3 and 4 were also able to test for two other types of synchronic bias. The first is another substantive bias, this one voicing-based rather than position-based. Because the experimental design included both neutralizing-to-voiced and neutralizing-to-voiceless conditions, participants could extend the stop voicing contrast positionally either by extending from voiced to voiceless stops or from voiceless to voiced stops, depending on the condition.

Because voiced stops are more marked than voiceless stops (in particular, they are hard to produce in word-edge position), it is more natural to extend from voiced stops to voiceless stops than vice versa. Thus the prediction from substantive bias was that, within initial and final contrast conditions, there would be greater extension of the stop voicing contrast in neutralizing-to-voiced conditions, in which participants had to extend to voiceless stops, than in neutralizing-to-voiceless conditions, in which participants had to extend to voiced stops.

The second additional type of bias was a complexity bias. Experiments 3 and 4 could test for complexity bias due to the nature of the filler consonants in the stimuli and, again, because there were both neutralizing-to-voiced and neutralizing-to-voiceless conditions. In Experiment 3, the filler consonants were sonorants, so in neutralizing-to-voiced conditions all segments in the word-edge position without the stop voicing contrast were [+voice] and segments not heard in this position during exposure could be excluded with the constraint *[–voice]. In neutralizing-to-voiceless conditions, the constraint *[–son, +voice] was needed to exclude segments not heard in the word-edge position without the stop voicing contrast. In Experiment 4, the filler consonants were voiceless fricatives, so in neutralizing-to-voiceless conditions all segments in the word-edge position without the stop voicing contrast were [–voice] and segments not heard in this position during exposure could be excluded with the constraint *[+voice]. In neutralizing-to-voiced conditions, the necessary phonotactic constraint was *[–cont, –voice]. In each experiment, then, one pair of conditions was more complex than the other pair and therefore expected to be more difficult, leading to greater extension of the stop voicing contrast (i.e. greater erroneous acceptance of test items of the type not heard in training). Concretely, the prediction from complexity bias was greater extension in neutralizing-to-voiceless conditions in Experiment 3 and greater extension in neutralizing-to-voiced conditions in Experiment 4. Notice that the predictions from voicing-based substantive bias and complexity bias are identical for Experiment 4 but opposite for Experiment 3.

In Experiment 3, participants extended the voicing contrast more in neutralizing-to-voiceless conditions than in neutralizing-to-voiced conditions. This effect was significant, and it supports complexity bias, not voicing-based substantive bias. In Experiment 4, participants extended the voicing contrast more in neutralizing-to-voiced conditions than in neutralizing-to-voiceless conditions; this effect was marginally significant. It fulfills the predictions of both voicing-based substantive bias and complexity bias, but since Experiment 3 could distinguish the two predictions and supported only complexity bias, I interpreted the results of Experiment 4 as supporting complexity bias as well. Thus in addition to providing some evidence of (position-based) substantive bias, the second set of phonotactic learning experiments yielded solid support for complexity bias.

The final experiment, Experiment 5, was the sole alternation learning experiment but partially resembled Experiment 3 in its design and predictions. This experiment tested the learnability of three types of stem-final stop voicing alternations: word-final devoicing, word-final voicing, and an exchange pattern in which stem-final [α voice] stops became [−α voice] in word-final position. Like in Experiments 3 and 4, the prediction from voicing-based substantive bias was that final devoicing would be learned better than final voicing. As in Experiment 3, the filler stem-final consonants were sonorants, so in the final voicing language all word-final consonants were [+voice] while in the final devoicing language some word-final consonants were [+voice] and others [−voice]. The final voicing language was therefore simpler, able to be mastered with the constraint *[−voice]# while the final devoicing language required *[−son, +voice]# (analogous, respectively, to the neutralizing-to-voiced and neutralizing-to-voiceless final contrast conditions in Experiment 3). We also predicted that the exchange pattern would be learned worse than both final voicing and devoicing due to its more complex, circular alternation.

The results of Experiment 5 showed that participants learned to apply final voicing and the exchange pattern (the latter was a marginally significant effect) but not final devoicing. Performance in the final voicing condition was overall better than in the final devoicing condition, and performance on alternating stems was better in the exchange condition than in the final devoicing condition. Recalling Experiment 3, voicing-based substantive bias and complexity bias make opposite predictions about the relative learnability of final voicing and devoicing. The fact that final voicing was learned better than final devoicing supports complexity bias, not substantive bias. While the learning of the exchange pattern was surprisingly good, I argued that it was not truly inconsistent with complexity bias because participants in the exchange condition were exposed to more alternating stems than participants in the other two conditions. This richer exposure may have boosted their performance.

5.2 Implications

5.2.1 Complexity Bias

Synthesizing these results, one clear finding is that learners are better at acquiring phonological patterns that can be expressed with fewer features. All three experiments whose designs could test this question found an effect in the right direction, providing further evidence for complexity bias, a synchronic learning bias for which there is already ample support in the experimental literature (Moreton & Pater 2012a; see also Section 1.2.1). In my experiments, the simpler patterns are simpler by virtue of being definable by constraints with a smaller number of feature instances, rather than, say, by virtue of relating instances of the same feature. The more novel aspect of my complexity bias results is that they depend crucially on the features of the filler segments in the stimuli and therefore on the structure and phonemic inventory of the entire artificial language. In the past, AGL experiments testing explicitly for complexity bias have had no filler segments, typically because the patterns being compared are vowel-vowel dependencies and all vowels participate in the patterns (Pycha et al. 2003, Moreton 2008, Baer-Henney & van de Vijver 2012, though see Finley & Badecker 2009 for a more complex case). Additionally, filler segments have not always been taken into account in explaining AGL results, as with Greenwood's (2016) final devoicing experiment.

5.2.2 Substantive Bias

The substantive bias results are far less clear-cut and raise important questions. While all five experiments tested for substantive bias, only Experiments 3 and 4, the phonotactic learning experiments involving the stop voicing contrast, yielded any learnability differences consistent with a naturalness bias. Of these two experiments, only Experiment 4 showed a reliable, significant effect of substantive bias. Given this showing, it is natural to ask whether substantive bias exists at all. Moreton & Pater (2012a) were skeptical, given the evidence, or more accurately the lack of evidence, available at the time, and though some promising results have been obtained since (Shapp 2012, Finley 2012, White 2013, van de Vijver & Baer-Henney 2014), more confounded, problematic, or null results have also accumulated since (see Section 1.2.2). It is possible that there is no synchronic learning bias that favors natural phonological patterns and that the phonetic naturalness of phonological systems cross-linguistically is only the result of sound change (Ohala 1992, Hale & Reiss 2000, Blevins 2004). I am reluctant to dismiss the existence of substantive bias so completely, though, precisely because of the few notable results in the literature, including the one in Experiment 4. These results are admittedly rare, suggesting that substantive bias is subtle and therefore often overwhelmed or washed out, but it is difficult to explain them away entirely.

Assuming, then, that the effects of substantive bias found through experimentation reflect something real about phonological learning, the question is why substantive bias emerges in some cases and not in others. In the context of this dissertation's experiments, this question can be broken down into more specific questions. First, why was there an effect of substantive bias in the positional extension of the stop voicing contrast but not of the major place contrasts in stops? The phonetic motivation for the word-final contrast entailing the word-initial contrast, but not vice versa, is there for both stop voicing and place of articulation, but participants were never more inclined to extend the three-way place contrast from word-final to word-initial position than from word-initial to word-final position. The difference between the place contrasts experiments and the voicing contrast experiments suggests that the participants were less sensitive to or had less phonetic knowledge of the weaker perceptibility of major place contrast in word-final position than the weaker perceptibility of the stop voicing contrast in word-final position. I have not found a satisfactory explanation for the differing reproducibility of the place contrasts implicational and the voicing contrast implicational in my phonotactic learning experiments. It remains a puzzle.

The second more specific question is why in Experiment 4 the perception-driven position-based substantive bias emerged but in Experiments 3 and 5 the articulation-driven voicing-based substantive bias did not emerge. Recall that the position-based substantive bias tested in Experiment 1-4 was an expected naturalness effect rooted in perceptual ease. Place and voicing contrasts in stops are more perceptible word-initially than word-finally. If learners possess this perceptual knowledge, they should infer such contrasts exist word-initially if they have heard them word-finally. The voicing-based substantive bias that Experiments 3-5 could test for was an expected naturalness effect rooted in articulatory ease. Voiced stops require more effort to produce at word edges than voiceless stops. Learners should therefore assume voiceless stops exist in a given word-edge position if they have heard voiced stops in that position (Experiments 3 and 4) or prefer a pattern in which voiced stops devoice word-finally than one in which voiceless stops voice word-finally (Experiment 5). In Experiment 4, the voicing-based substantive bias was confounded with complexity bias, and I attributed the results consistent with voicing-based substantive bias entirely to complexity bias. This case aside, Experiments 3-5 yielded some evidence for the perceptual position-based substantive bias but never any evidence for the articulatory voicing-based substantive bias.

Interference from another type of bias may explain the apparent lack of any effect of voicing-based substantive bias in this dissertation. In Experiments 3 and 5, complexity bias and voicing-based substantive bias would have pushed learners in opposite directions, and the greater strength of complexity bias, suggested by the literature, may have caused it to prevail over the weaker voicing-based substantive bias. That is, any effect the articulatorily-rooted substantive bias may have had in these experiments could have been obscured by a stronger bias counteracting the substantive bias and not only nullifying it but overwhelming it to yield the opposite effect. Voicing-based substantive bias may simply not have had the opportunity to emerge. That said, the difference between perceptually-rooted and articulatorily-rooted substantive bias deserves more attention.

5.2.3 Perceptually-Grounded vs. Articulatorily-Grounded Substantive Bias

The patchwork of positive and null results that plagues the experimental substantive bias literature may be a consequence of treating substantive bias as a single, undifferentiated constraint on learning when it should be broken down into subtypes for which there may be more or less evidence. Specifically, I propose to distinguish between perceptual substantive bias and articulatory substantive bias, and I claim that only perceptual substantive bias constrains phonological learning. This claim was first made in Glewwe et al. 2018 and grew out of collaborative work; here I develop it further within the context of my own outlook on substantive bias.[43]

The only substantive bias effects found in the experiments in this dissertation are rooted in perception; effects rooted in articulation failed to come out. While in these experiments that may have been due to the obscuring effects of other biases, the wider AGL literature also suggests that articulatory ease does not synchronically shape phonological learning while perceptual advantages can. Before assessing the literature, I offer a few cautions. First, AGL experiments of course vary greatly along a number of dimensions, many of which relate to their design, so a single positive or null result may be due to any number of factors or combinations of factors. Second, it is not always perfectly clear whether the phonetic naturalness of a phonological pattern is articulatorily-grounded or perceptually-grounded. Articulation and perception are not divorced from each other; a pattern that facilitates articulation may also result in perceptual benefits. A good example is vowel harmony, which promotes ease of articulation by reducing the number of gestures needed for production but also results in vowels that provide

---

[43] Thanks to Jesse Zymet, who was the first to conjecture that only perceptual factors, and not articulatory ease, bias phonological learning.

cues to the quality of neighboring vowels. I discussed the phonetic motivation for vowel harmony in more detail in Section 1.2.2.2. For present purposes, I consider the broad phenomenon of vowel harmony to be articulatorily motivated while asymmetries within the typology of vowel harmony (e.g. properties of triggers, properties of targets) may be perceptually motivated. This seems to be how vowel harmony has been treated in the AGL literature. More generally, AGL studies usually explain the phonetic motivation for the natural pattern being tested, appealing to either articulation or perception. For the most part, I classify experiments by the type of motivation the authors invoked. I exclude from the following discussion purported naturalness results that can be reinterpreted as effects of complexity bias (see Section 1.2.2.3).

First consider the studies whose results provided more or less convincing evidence for substantive bias (see Sections 1.2.2.4 and especially 1.2.2.5 for details and occasional critiques of the substantive bias interpretations). Wilson (2006) found that alternations relating velar stops and palato-alveolar affricates were extended to prevocalic environments in which the stops and affricates were more perceptually similar but not to environments in which they were less similar. White (2013) also found that learners preferred for more perceptually similar sounds to alternate with each other and tended to infer that more perceptually similar sounds alternated when taught that perceptually distant sounds did. In Carpenter’s (2006) stress and vowel height experiments, English and French speakers learned a perceptually natural pattern in which low vowels attracted stress better than an unnatural pattern in which high vowels attracted stress. Mid vowels derive a greater perceptual benefit from triggering rounding harmony than high vowels. Participants learn mid vowel-triggered rounding harmony better than high vowel-triggered rounding harmony (Finley 2012) and extend rounding harmony from high triggers to mid triggers more than from mid triggers to high triggers (Kimper 2016). Similarly, front vowels are more often targets of height harmony than back vowels, and one possible motivation for this asymmetry is perceptual: height harmony may hurt the perceptibility of back vowels. In an experiment, front vowel targets of height harmony were learned better than back vowel targets, and learners extended height harmony from back targets to front targets but not vice versa (Finley & Badecker 2012). Finally, Greenwood (2016) found better learning a perceptually natural pattern in which heavy syllables attracted stress than an unnatural pattern in which light syllables attracted stress, but according to a perception test, stress was perceived less accurately in the unnatural condition, so this result may be due to in-experiment channel bias rather than participants' perceptual knowledge independent of the experiment. All these substantive bias effects rely on phonetic knowledge of relative perceptual similarity and salience.

Other positive substantive bias results come from Shapp 2012 and Baer-Henney 2015. Shapp argued that sibilant harmony is more natural than voicing harmony insofar as it promotes articulatory ease in a way that voicing harmony does not, and she found superior learning of sibilant harmony over voicing harmony. Baer-Henney tested the learning of intervocalic stop voicing at different places of articulation since the articulatory difficulty of voicing a stop increases as place of articulation becomes more posterior. She found that intervocalic voicing of dorsal stops was learned worse than intervocalic voicing of labial and coronal stops. This result was consistent with an effect of articulatory substantive bias, but the degrees to which participants trained on voicing at a particular place of articulation extended to the other places of articulation were not uniformly so.

Though the number of studies is admittedly not large, the great majority of AGL experiments yielding promising evidence for substantive bias test for naturalness rooted in perceptual factors. The naturalness effect found in Experiment 4 of this dissertation adds to this preponderance of perceptually-driven substantive bias effects.

Turning to null substantive bias results, the imbalance goes the other way: most of the studies that have failed to find evidence for substantive bias tested for superior learning of a pattern that was argued to be natural on articulatory grounds. There is the case of vowel harmony, which was briefly discussed above: participants do not learn vowel harmony better than vowel disharmony (Pycha et al. 2003, Skoruppa & Peperkamp 2011). Another case that has been tested multiple times is a pattern in which all coda/final obstruents are voiceless vs. a pattern in which all coda/final obstruents are voiced. The first pattern is articulatorily more natural, but it is learned no better than the second pattern (Saffran & Thiessen 2003, Myers & Padgett 2014, Greenwood 2016 Chapter 6). To be fair, the effects of articulatorily-driven substantive bias in Myers & Padgett 2014 and Greenwood 2016 may have been balanced out (Myers & Padgett) or overridden (Greenwood) by an effect of complexity bias, as in my Experiments 3 and 5. Other experiments have also tested for substantive biases related to voicing. Peperkamp & Dupoux (2007) compared the learnability of intervocalic voicing, an articulatorily natural alternation, and an arbitrary alternation and found no difference. Do, Zsiga, & Havenhill (2016) tested the learning of articulatorily natural postnasal voicing and articulatorily unnatural postnasal devoicing and also found no difference. Lysvik (2018) compared the learnability of articulatorily natural final devoicing and articulatorily unnatural final voicing. Unlike in our Experiment 5, there was no way in which one pattern was simpler than the other, and both were learned to the same degree. Finally, Seidl & Buckley (2005) tested infants’ learning of natural and unnatural or arbitrary phonotactic generalizations. Their first experiment compared the learnability of a phonotactic system consistent with intervocalic lenition of stops to affricates/fricatives, an articulatorily motivated pattern, and an articulatorily unnatural phonotactic system in which affricates/fricatives could only be word-initial while stops could be intervocalic. Infants learned both patterns equally well. The second experiment compared a pattern in which labial consonants were followed by round vowels and coronal consonants by front vowels, both dependencies that arguably promote articulatory ease, and a minimally different pattern with arbitrary consonant-vowel dependencies. Again, infants learned both patterns equally well.

All the studies just reviewed tested for better learning of articulatorily-grounded patterns over patterns lacking articulatory grounding, and all of them turned up null results. None of these studies used a production task, and one might be tempted to attribute the lack of learnability differences to this fact. However, Shapp 2012 and Baer-Henney's (2015) intervocalic voicing experiment did not use production tasks either and still found effects of apparently articulatorily-rooted substantive bias. (Additionally, it is not clear how production tasks testing for articulation-based substantive bias could get away from Greenwood's (2016) objection to in-experiment channel bias since participants would be experiencing varying degrees of articulatory difficulty in the course of the experiment. This is a question for a separate discussion, though.)

Null results in studies testing for effects of perception-based substantive bias are less common. Coming back to rounding harmony, the greater perceptibility of roundedness in high vowels makes them better targets of rounding harmony than mid vowels because high vowels can better signal the roundedness of the triggering vowel. Despite this, participants extended as much from high targets of rounding harmony to mid targets as from mid targets to high targets (Finley & Badecker 2009 Experiment 3, Finley 2008 Experiment 7). In Carpenter's (2006) stress and syllable weight experiments, a perceptually natural pattern in which heavy syllables attracted stress was learned no better than an unnatural pattern in which light syllables attracted stress, though Carpenter raised an issue with her stimuli that may have led to this null result. Finally, Greenwood's (2016) study of coda sonorancy found equal learning of a sonorant codas only pattern and a voiceless stop codas only pattern even though she argued that sonorants' greater perceptual salience made them more natural codas.

Drawing these studies together, it emerges that the majority of null substantive bias results are cases in which the natural pattern was natural for articulatory reasons. Looking at this body of AGL studies another way, experiments that test for perception-based substantive bias are more likely to find an effect than not while experiments that test for articulation-based substantive bias are more likely to yield a null result than not. The generalizations are not exceptionless, but they are suggestive. It is therefore necessary, when evaluating the substantive bias literature, to distinguish between perceptual and articulatory grounding. Our understanding of substantive bias should be more fine-grained.

Based on the experiments in this dissertation and the wider literature, I also believe perceptual substantive bias holds more promise than articulatory substantive bias. This is a claim based on experimental evidence. It immediately raises the question of *why* only perceptual naturalness and not articulatory naturalness should bias phonological learning. I cannot yet propose an answer to this question, and so here I offer only some very preliminary speculations. In phonological acquisition, production lags behind perception, both for children learning their native language (Braine 1974:284, Hayes 1999:266 and references therein) and for second-language learners (Broselow 2018). It seems conceivable that the reason perceptual factors but not articulatory factors constrain phonological learning might relate to perception's earlier development.[44] Elaborating a theory of such a relation, however, would require much further investigation.

Another possibility is that Steriade's (2001b) P(erceptual)-map exists but there is no corresponding A(rticulatory)-map. That is, there is a component of the grammar that stores knowledge of perceptual similarity that can then influence what kinds of phonological patterns learners prefer, but there is no equivalent component of the grammar that stores knowledge of articulatory ease that could then influence phonological learning. Hayes (1999) has in fact proposed the existence of phonetic maps that encode articulatory information, and Byun, Inkelas, & Rose (2016) have proposed a theory of the A(rticulatory)-map to account for child phonology, though their A-map is substantially different from the phonetic maps Steriade and Hayes envision. It would be difficult to prove the non-existence of an articulatory analogue to the P-map, and even if one were to make such a proposal, the question of why there should be a P-map but no A-map would remain.

Clearly there are still many questions to address. Future research should break down substantive bias into narrower types, carefully investigate exactly what articulatory and/or perceptual considerations motivate particular phonological patterns deemed phonetically natural, and further test the hypothesis I have put forth in this section.

5.3 Future Directions

Future experiments could confirm some of the effects I have argued for in the interpretation of the present studies. I designed Experiment 4 to test the validity of the complexity bias account of Experiment 3. This was accomplished by changing the voicing of the

---

[44] I thank Bruce Hayes for pointing this out.

filler consonants from voiced in Experiment 3 to voiceless in Experiment 4. However, since the filler consonants in Experiment 3 were sonorants, making them voiceless in Experiment 4 also required changing their sonorancy. The complexity bias I found in Experiment 3 also emerged in Experiment 4, but only marginally so, and I suggested this was because voiced counterparts to the voiceless filler consonants exist in the phoneme inventory of English, the participants' native language. Another follow-up that would be a closer counterpart to Experiment 4 than Experiment 3 is an analogous experiment in which the filler consonants are voiced but are voiced fricatives instead of the voiced sonorants of Experiment 3. Such an experiment would be the true minimally different complementary study to Experiment 4. The complexity bias predictions would be the same as for Experiment 3, but the sonorancy of the filler consonants would no longer entail their voicedness, a fact which perhaps gave the participants in Experiment 3 an advantage.[45]

Similarly, follow-up studies to Experiment 5 could replace the filler consonants with voiced fricatives in one version and voiceless fricatives in another, again to test for effects of complexity bias. Making filler consonants differ in only a single feature limits the availability of alternative explanations for a difference in results depending on filler consonants. In the voicing alternations learning study, I predict that voiced fricative filler consonants should lead to better learning of final voicing than of final devoicing, as we found in Experiment 5. Conversely, voiceless fricative filler consonants should lead to better learning of final devoicing than of final voicing.

---

[45] It may seem more probable for participants to misperceive voiceless fricatives as devoiced productions of underlyingly voiced fricatives (as was possible in Experiment 4) than for them to misperceive fully voiced fricatives in experimental stimuli as underlyingly voiceless fricatives. This would mean misperception is still more likely in a version of the voicing contrast experiment where filler consonants are voiceless fricatives than in a version where filler consonants are voiced fricatives. The fact remains, however, that misperception by English speakers is possible with (even hyperarticulated) voiced fricatives in a way that it is not for voiced sonorants.

The complexity bias effects I have argued for in Experiments 3-5 are all rooted in the sound inventories of the artificial languages to which learners were exposed. To further test my hypotheses about the role of filler consonants in AGL experimentation and the interpretations of the results in this dissertation, I could run analogues of Experiment 3 and Experiment 5 in which both voiced and voiceless sounds (e.g. voiced and voiceless fricatives) served as filler consonants. As mentioned in Section 3.2.4, this design would eliminate all predictions from complexity bias because participants would be exposed to voiced and voiceless segments in all word-edge positions. It would no longer be possible to rely on a single-feature constraint to exclude all segments with a particular value for [voice] in a given position. In every condition, participants would have to learn a restriction specifically against stops with a particular voicing value, and the simplest constraint needed to master the artificial language would have the same level of featural complexity. I would expect equal learning in all conditions unless substantive biases are at work. In particular, in this design the articulatorily-rooted, voicing-based substantive bias would no longer be confounded with complexity bias, as was the case in Experiment 4. Nor would it have to overcome complexity bias in order to emerge, as in Experiments 3 and 5. If in this design there was still no effect of voicing-based substantive bias, it would further bolster the claim that articulatory substantive bias does not constrain phonological learning.

Another question future research could try to address is whether substantive bias is innate or learned. This is likely to be a difficult question to test, but cross-linguistic investigations may be able to shed some light on the issue. Thanks to a shared articulatory apparatus and auditory system, speakers of all languages will have many similar phonetic experiences (e.g. articulatory difficulties, perceptual effects). There are universal aspects of production and perception that, for evolutionary phonologists and others with similar views, are precisely what give rise to strong tendencies and near-universals in the phonological typology (Ohala 1974, 1992, Blevins 2004). This makes it difficult to categorize phonetic knowledge as innate or learned because most speakers, of any language, have had the opportunity to acquire the same phonetic knowledge from their linguistic experience. Thus the ubiquity of some type of phonetic knowledge does not necessarily mean that it is innate. Nevertheless, I believe there is scope for demonstrating that some phonetic knowledge is learned. Different languages may still provide their speakers with unequal opportunities to acquire certain pockets of phonetic knowledge. As White (2013:167) suggested, the perceptual knowledge that makes up the P-map may be learned, not innate, which opens up the possibility of the P-map being language-dependent (a possibility Steriade (2001b:44) herself leaves room for). If the P-map varies from language to language, speakers of different languages may have perceptual knowledge that constrains or does not constrain their phonological learning in different ways. If an AGL experiment captured such differences in learning, for instance, the presence vs. absence of a bias effect, between participant populations with different native languages, this would suggest that the relevant perceptual knowledge was learned for the speakers who exhibited the effect.

To propose a concrete test, consider my Experiments 3 and 4, where I predicted participants would use their knowledge of the relative perceptual similarity of voiced stops to voiceless stops in different positions to infer a word-initial voicing contrast when exposed to a word-final contrast. What makes voiced and voiceless stops more perceptually similar in word-final position in English is the absence of release cues, which help distinguish voicing in word-initial position. Some languages, though, including French, rely mostly on closure cues rather than release cues to distinguish voicing in stops (Blevins 2004). Closure cues are available both word-initially and word-finally, so it is conceivable that French speakers do not acquire the same perceptual knowledge about the greater similarity of voiced and voiceless stops in word-final position because this greater similarity is not a property of their native language input. If this is the case, French speakers should perform differently in Experiments 3 and 4; they should not extend the stop voicing contrast more from word-final to word-initial position than from word-initial to word-final position. If French speakers indeed behave differently, this would support the hypothesis that the perceptual knowledge driving English speakers' asymmetrical extension of the stop voicing contrast is learned, not innate. In general, conducting the same AGL experiments on speakers of different languages could yield important insights into the language-specificity or universality of biases rooted in phonetic naturalness.

Appendix: Experiment Stimuli

For stimuli in Experiments 1-4: Shaded items are familiar (heard in training and test). Bolded items belong to minimal pairs or triplets.

Experiment 1 Stimuli: Place Contrast

	#P	#T	#K	P#	T#	K#
Training	**pínir**	**tínir**	**kínir**	**jáwip**	**jáwit**	**jáwik**
	pilán	tajál	kilún	rujáp	wirút	wanúk
	pulír	**tulír**	**kulír**	**luríp**	**lurít**	**lurík**
	pírul	túwan	kíral	júnup	lánat	nálak
	bínar	**dínar**	**gínar**	**nílab**	**nílad**	**nílag**
	bilún	dirúl	gurúl	lijúb	julád	jalág
	bawán	**dawán**	**gawán**	**rujúb**	**rujúd**	**rujúg**
	búnil	dánil	gírul	jáwab	nírad	línig
		tíjar			wíjat	
		tinúl			nilút	
		talín			wajút	
		tújal			rúwat	
		dánir			wínud	
		dalúr			nalíd	
		duján			jarúd	
		díwil			lúnid	
	#P	#T	#K	P#	T#	K#
Test	**pínir**	**tínir**	**kínir**	**nilíp**	**nilít**	**nilík**
	pilán	tajál	kilún	nálup	nájat	líruk
	panúl	**tanúl**	**kanúl**	**jáwip**	**jáwit**	**jáwik**
	páwir	túrul	kájar	ruják	wirút	wanúk
	bawán	**dawán**	**gawán**	**wánub**	**wánud**	**wánug**
	búnil	dánil	gírul	waríb	luwíd	nurág
	búlar	**dúlar**	**gúlar**	**rujúb**	**rujúd**	**rujúg**
	burún	dajún	galín	jáwab	nírad	línig

Experiment 2 Stimuli: Place Contrast with Fillers

	#P	#T	#K	P#	T#	K#	Fillers
Training	**pínir**	**tínir**	**kínir**	**jáwip**	**jáwit**	**jáwik**	níjal
	pilán	tajál	kilún	rujáp	wirút	wanúk	rílun
	pulír	**tulír**	**kulír**	**luríp**	**lurít**	**lurík**	jáwar
	pírul	túwan	kíral	júnup	lánat	nálak	rújun
	bínar	**dínar**	**gínar**	**nílab**	**nílad**	**nílag**	líwir
	bilún	dirúl	gurúl	lijúb	julád	jalág	náwal
	bawán	**dawán**	**gawán**	**rujúb**	**rujúd**	**rujúg**	júnar
	búnil	dánil	gírul	jáwab	nírad	línig	wírin
							lánir
		tíjar			wíjat		járan
		tinúl			nilút		lirín
		talín			wajút		wijúl
		tújal			rúwat		janír
		dánir			wínud		ralún
		dalúr			nalíd		luwín
		duján			jarúd		nirál
		díwil			lúnid		rijún
							wanúl
							liwár
							warín
Test	#P	#T	#K	P#	T#	K#	
	pínir	**tínir**	**kínir**	**nilíp**	**nilít**	**nilík**	
	pilán	tajál	kilún	nálup	nájat	líruk	
	panúl	**tanúl**	**kanúl**	**jáwip**	**jáwit**	**jáwik**	
	páwir	túrul	kájar	ruják	wirút	wanúk	
	bawán	**dawán**	**gawán**	**wánub**	**wánud**	**wánug**	
	búnil	dánil	gírul	waríb	luwíd	nurág	
	búlar	**dúlar**	**gúlar**	**rujúb**	**rujúd**	**rujúg**	
	burún	dajún	galín	jáwab	nírad	línig	

Experiment 3 Stimuli: Voicing Contrast with Sonorant Filler Consonants

	…T#	…D#	#T…	#D…
Training	**míwip**	**míwib**	**pímir**	**bímir**
	nimáp	**nimáb**	**pinál**	**binál**
	lanít	**laníd**	**tumál**	**dumál**
	járat	**járad**	**túwim**	**dúwim**
	rújuk	**rújug**	**kájun**	**gájun**
	nuwák	**nuwág**	**kawám**	**gawám**
	lárip	láwib	páwin	bálin
	wajúp	waját	paját	bajúm
	níwit	mírid	tíril	dílur
	wimát	rilúd	tilár	dirín
	murúk	lunág	kuníl	gunál
	júnak	júmig	kúlan	gúrim
Test	…T#	…D#	#T…	#D…
	nimáp	**nimáb**	**pinál**	**binál**
	rínup	**rínub**	**pírum**	**bírum**
	járat	**járad**	**túwim**	**dúwim**
	walút	**walúd**	**tunúr**	**dunúr**
	nuwák	**nuwág**	**kawám**	**gawám**
	múlik	**múlig**	**kámir**	**gámir**
	lárip	láwib	páwin	bálin
	185alap	wamúb	palún	bawán
	níwit	mírid	tíril	dílur
	rijút	niját	tijún	diját
	murúk	lunág	kuníl	gunál
	lúmik	júrug	kúrum	gúlun

Experiment 4 Stimuli: Voicing Contrast with Fricative Filler Consonants

	…T#	…D#	#T…	#D…
Training	**físip**	**físib**	**pífis**	**bífis**
	θifáp	**θifáb**	**piθáʃ**	**biθáʃ**
	ʃaθít	**ʃaθíd**	**tufáʃ**	**dufáʃ**
	hásat	**hásad**	**túsif**	**dúsif**
	súfuk	**súfug**	**káʃuθ**	**gáʃuθ**
	θuʃák	**θuʃág**	**kaʃáf**	**gaʃáf**
	ʃásip	ʃáfib	páfiθ	báʃiθ
	hafúp	hafáb	pasáf	baʃúf
	θísit	físid	tísiʃ	díʃus
	sifát	siʃúd	tiʃás	disíθ
	fusúk	ʃuθág	kuθíʃ	guθáʃ
	húθak	húfig	kúʃaθ	gúsif
Test	…T#	…D#	#T…	#D…
	θifáp	**θifáb**	**piθáʃ**	**biθáʃ**
	síθup	**síθub**	**písuf**	**bísuf**
	hásat	**hásad**	**túsif**	**dúsif**
	faʃút	**faʃúd**	**tuθús**	**duθús**
	θuʃák	**θuʃág**	**kaʃáf**	**gaʃáf**
	fúʃik	**fúʃig**	**káfis**	**gáfis**
	ʃásip	ʃáfib	páfiθ	báʃiθ
	saʃáp	safúb	paʃúθ	bafáθ
	θísit	físid	tísiʃ	díʃus
	hifút	θisád	tisúθ	disáʃ
	fusúk	ʃuθág	kuθíʃ	guθáʃ
	ʃúfik	húsug	kúsuf	gúʃuθ

Experiment 5 Stimuli: Voicing Alternations

PL = plural; NON-ALT-SG = non-alternating singular; ALT-SG = alternating singular. Shaded plural items are familiar (heard in training and test). In the Devoicing condition, participants are trained on non-alternating singulars for voiceless-final stems and alternating singulars for voiced-final stems. In the Voicing condition, participants are trained on alternating singulars for voiceless-final stems and non-alternating singulars for voiced-final stems. In the Exchange condition, participants are trained on alternating singulars for voiceless-final and voiced-final stems.

Training	Sonorant-Final Stem			Voiceless-Final Stem			Voiced-Final Stem		
	PL	NON-ALT-SG	ALT-SG	PL	NON-ALT-SG	ALT-SG	PL	NON-ALT-SG	ALT-SG
	gipál-i	gipál		malót-i	malót	malód	rɛkɛ́b-i	rɛkɛ́b	rɛkɛ́p
	tikár-i	tikár		dibák-i	dibák	dibág	binóg-i	binóg	binók
	nigúm-i	nigúm		kinɛ́p-i	kinɛ́p	kinɛ́b	lomád-i	lomád	lomát
	podán-i	podán		rogɛ́p-i	rogɛ́p	rogɛ́b	pirɛ́d-i	pirɛ́d	pirɛ́t
	madúr-i	madúr		kɛnɛ́t-i	kɛnɛ́t	kɛnɛ́d	pɛnág-i	pɛnág	pɛnák
	gamɛ́l-i	gamɛ́l		labók-i	labók	labóg	romób-i	romób	romóp
	dabór-i	dabór		gadáp-i	gadáp	gadáb	kɛpɛ́d-i	kɛpɛ́d	kɛpɛ́t
	liróm-i	liróm		tonɛ́t-i	tonɛ́t	tonɛ́d	natóg-i	natóg	natók
	kodún-i	kodún		tɛmúk-i	tɛmúk	tɛmúg	bunáb-i	bunáb	bunáp
	ranúl-i	ranúl		rilát-i	rilát	rilád	nɛmúd-i	nɛmúd	nɛmút
	kidɛ́m-i	kidɛ́m		pɛbúk-i	pɛbúk	pɛbúg	rɛdúg-i	rɛdúg	rɛdúk
	pukár-i	pukár		galúp-i	galúp	galúb	butób-i	butób	butóp
Test	Sonorant-Final Stem			Voiceless-Final Stem			Voiced-Final Stem		
	PL	NON-ALT-SG	ALT-SG	PL	NON-ALT-SG	ALT-SG	PL	NON-ALT-SG	ALT-SG
	gipál-i	gipál	gipár	malót-i	malót	malód	rɛkɛ́b-i	rɛkɛ́b	rɛkɛ́p
	tikár-i	tikár	tikál	dibák-i	dibák	dibág	binóg-i	binóg	binók
	nigúm-i	nigúm	nigún	kinɛ́p-i	kinɛ́p	kinɛ́b	lomád-i	lomád	lomát
	podán-i	podán	podám	rogɛ́p-i	rogɛ́p	rogɛ́b	pirɛ́d-i	pirɛ́d	pirɛ́t
	madúr-i	madúr	madúl	kɛnɛ́t-i	kɛnɛ́t	kɛnɛ́d	pɛnág-i	pɛnág	pɛnák
	gamɛ́l-i	gamɛ́l	gamɛ́r	labók-i	labók	labóg	romób-i	romób	romóp
	timɛ́r-i	timɛ́r	timɛ́l	pogɛ́t-i	pogɛ́t	pogɛ́d	nɛpód-i	nɛpód	nɛpót
	mulɛ́n-i	mulɛ́n	mulɛ́m	mogáp-i	mogáp	mogáb	dirúg-i	dirúg	dirúk
	nɛpɛ́n-i	nɛpɛ́n	nɛpɛ́m	tilók-i	tilók	tilóg	romáb-i	romáb	romáp
	gonúl-i	gonúl	gonúr	kɛták-i	kɛták	kɛtág	bikɛ́b-i	bikɛ́b	bikɛ́p
	kapɛ́r-i	kapɛ́r	kapɛ́l	mudáp-i	mudáp	mudáb	pɛkúd-i	pɛkúd	pɛkút
	lipám-i	lipám	lipán	gapót-i	gapót	gapód	lɛtóg-i	lɛtóg	lɛtók

References

Anderson, S. & Browne, W. (1973). On keeping exchange rules in Czech. *Papers in Linguistics*, *6*, 445–482.

Baer-Henney, D. (2015). *Learners' Little Helper: Strength and Weakness of the Substantive Bias in Phonological Acquisition*. Ph.D. dissertation. University of Potsdam.

Baer-Henney, D., & van de Vijver, R. (2012). On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology*, *3*(2), 221–250.

Baer-Henney, D., Kügler, F., & van de Vijver, R. (2015). The Interaction of Language-Specific and Universal Factors During the Acquisition of Morphophonemic Alternations with Exceptions. *Cognitive Science*, *39*, 1537–1569.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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

Becker, M., Clemens, L. E., & Nevins, A. (2017). Generalization of French and Portuguese plural alternations and initial syllable protection. *Natural Language & Linguistic Theory*, *35*(2), 299–345.

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

Beguš, G. (2018). Bootstrapping Sound Changes. Ms. https://ling.auf.net/lingbuzz/004299.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Bernhardt, B. H. & Stemberger, J. P. (1998). Acquisition of Adult Alternations. In B. H. Bernhardt & J. P. Stemberger (Eds.), *Handbook of Phonological Development: From the Perspective of Constraint-Based Nonlinear Phonology* (pp. 635–666). San Diego: Academic Press.

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

Braine, M. D. S. (1974). On What Might Constitute Learnable Phonology. *Language*, *50*(2), 270–299.

Brockhaus, W. (1995). *Final Devoicing in the Phonology of German*. Tübingen: Niemeyer.

Broselow, E. (2018). Laryngeal contrasts in second language phonology. In L. Hyman & F. Plank (Eds.), *Phonetics and Phonology 23: Phonological Typology* (pp. 312–340). Berlin/Boston: Walter de Gruyter.

Bye, P. (2011). Dissimilation. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology, Volume 3, Phonological Processes* (pp. 1408–1433). Oxford: Wiley-Blackwell.

Byun, T. M., Inkelas, S., & Rose, Y. (2016). The A-Map Model: Articulatory Reliability in Child-Specific Phonology. *Language*, *92*(1), 141–178.

Carpenter, A. C. (2006). Acquisition of a Natural Versus an Unnatural Stress System. Ph.D. dissertation. University of Massachusetts Amherst.

Carpenter, A. C. (2010). A naturalness bias in learning stress. *Phonology*, *27*(3), 345–392.

Chomsky, N. & Halle, M. (1965). Some controversial questions in phonological theory. *Journal of Linguistics*, *1*(2), 97–138.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Chong, A. J. (2017). *On the Relation between Phonotactic Learning and Alternation Learning*. Ph.D. dissertation. University of California, Los Angeles.

Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, *170*, 312–327.

Cristià, A., & Seidl, A. (2008). Is Infants' Learning of Sound Patterns Constrained by Phonological Features? *Language, Learning, and Development*, *4*(3), 203–227.

de Lacy, Paul. (2006). *Markedness: Reduction and Preservation in Phonology*. Cambridge: Cambridge University Press.

Do, Y. (2018). Paradigm uniformity bias in the learning of Korean verbal inflections. *Phonology*, *35*(4), 547–575.

Do, Y., Zsiga, E. & Havenhill, J. (2016). Naturalness and frequency in implicit phonological learning. Talk presented at the *90th Annual Meeting of the Linguistic Society of America*, Washington, D.C.

Durvasula, K. & Nelson, S. (2018). Lexical Returning Targets Features. In G. Gallagher, M. Gouskova, & S. Yin (Eds.), *Proceedings of the 2017 Annual Meeting on Phonology*. Washington, D.C.: Linguistic Society of America.

Finley, S. (2008). *Formal and Cognitive Restrictions on Vowel Harmony*. Ph.D. dissertation. Johns Hopkins University.

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

Finley, S. & Badecker, W. (2008). Analytic Biases for Vowel Harmony Languages. In N. Abner & J. Bishop (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics (WCCFL 27)* (pp. 168–176). Somerville, MA: Cascadilla Press.

Finley, S. & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, *61*(3), 423–437.

Finley, S. & Badecker, W. (2012). Learning Biases for Vowel Height Harmony. *Journal of Cognitive Science*, *13*, 287–327.

Frisch, S. A. (2004). Language processing and segmental OCP effects. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 346–371). Cambridge: Cambridge University Press.

Frisch, S. A. & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*(2), 139–162.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*(1), 27–52.

Fujimura, O., Macchi, M. & Streeter, L. A. (1978). Perception of Stop Consonants with Conflicting Transitional Cues: A Cross-Linguistic Study. *Language and Speech*, *21*(4), 337–346.

Glewwe, E., Zymet, J., Adams, J., Jacobson, R., Yates, A., Zeng, A., & Daland, R. (2018). Substantive bias and the acquisition of final (de)voicing patterns. Talk presented at the *92nd Annual Meeting of the Linguistic Society of America*, Salt Lake City.

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

Goldwater, S. & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In J. Spenader, A. Eriksson, & Östen Dahl (Eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory* (pp. 111–120). Stockholm: Stockholm University, Department of Linguistics.

Gordon, M. K. (2016). *Phonological Typology*. Oxford: Oxford University Press.

Greenberg, J. H. (1978). Some Generalizations Concerning Initial and Final Consonant Clusters. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of Human Language, Volume 2, Phonology* (pp. 243–280). Stanford: Stanford University Press.

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

Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics, 6*(2), 95–102.

Hale, M. & Reiss, C. (2000). "Substance Abuse" and "Dysfunctionalism": Current Trends in Phonology. *Linguistic Inquiry*, *31*(1), 157–169.

Hale, M., Kissock, M., & Reiss, C. (2007). Microvariation, variation, and the features of universal grammar. *Lingua*, *117*(4), 645–665.

Hayes, B. (1999). Phonetically-Driven Phonology: The Role of Optimality Theory and Inductive Grounding. In M. Darnell, E. Moravscik, M. Noonan, F. Newmeyer, & K. Wheatly (Eds.), *Functionalism and Formalism in Linguistics*, *Volume I: General Papers* (pp. 243–285). Amsterdam: John Benjamins.

Hayes, B. & Steriade, D. (2004). Introduction: the phonetic bases of phonological Markedness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 1–33). Cambridge: Cambridge University Press.

Hayes, B. & Stivers, T. (1995). Postnasal Voicing. Ms. University of California, Los Angeles.

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

Hayes, B. & Wilson, C. (2008). A Maximum Entropy Model of Phonotactic and Phonotactic Learning. *Linguistic Inquiry*, 39, 379–440.

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

Johnson, L. (1973). Dissimilation as a Natural Process in Phonology. *Stanford Occasional Papers in Linguistics*, *3*, 45–56.

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

Kapatsinski, V. (2013). Conspiring to Mean: Experimental and Computational Evidence for a Usage-Based Harmonic Approach to Morphophonology. *Language*, *89*(1), 110–148.

Kaun, A. (2004). The typology of rounding harmony. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 87–116). Cambridge: Cambridge University Press.

Kazazis, K. (1969). Possible evidence for (near-)underlying forms in the speech of a child. In R. I. Binnick, A. Davidson, G. M. Green, & J. L. Morgan (Eds.), *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society* (pp. 382–388). Chicago: Department of Linguistics, University of Chicago.

Kean, M.-L. (1975). *The Theory of Markedness in Generative Grammar*. Ph.D. dissertation. Massachusetts Institute of Technology.

Kenstowicz, M. (1997). Quality-sensitive stress. *Rivista di Linguistica*, *9*(1), 157–187.

Kerkhoff, A. (2004). Acquisition of voicing alternations. In J. van Kempen & S. Baauw (Eds.), *Proceedings of GALA 2003 (Generative Approaches to Language Acquisition), Vol. 2* (pp. 269–280). Utrecht: LOT.

Kimper, W. (2016). Asymmetric Generalisation of Harmony Triggers. In G. Ó. Hansson, A. Farris-Trimble, K. McMullin, & D. Pulleyblank (Eds.), *Proceedings of the 2015 Annual Meeting on Phonology*. Washington, D.C.: Linguistic Society of America.

Legendre, G., Miyata, Y., & Smolensky, P. (1990). Harmonic Grammar — A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 388–395). Institute of Cognitive Science Technical Report #90-5. University of Colorado at Boulder.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Linzen, T. & Gallagher, G. (2017). Rapid generalization in phonotactic learning. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1), 24.

Lisker, L. & Abramson, A. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*, 1–28.

Lombardi, L. (1991). *Laryngeal features and laryngeal neutralization.* Ph.D. dissertation. University of Massachusetts.

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.

Lysvik, J. K. (2018). An Artificial Language Learning experiment finds no bias against word-final voicing. Poster presented at the *Twenty-Sixth Manchester Phonology Meeting*, Manchester.

Mielke, J. (2004). *The Emergence of Distinctive Features*. Ph.D. dissertation. The Ohio State University.

Moreton, E. (1999). Non-Computable Functions in Optimality Theory. ROA-364. Rutgers Optimality Archive. http://roa.rutgers.edu/.

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.

Moreton, E. & Pertsova, K. (2016). Implicit and Explicit Processes in Phonotactic Learning. In J. Scott & D. Waugtal (Eds.), *Proceedings of the 40th Boston University Conference on Language Acquisition (BUCLD 40)* (pp. 277–290). Somerville, MA: Cascadilla Press.

Moreton, E., Pater, J. & Pertsova, K. (2017). Phonological Concept Learning. *Cognitive Science, 41*, 4–69.

Mowrey, R. A. & MacKay, I. R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America, 88*(3), 1299–1312.

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

Ohala, J. J. (1974). Phonetic Explanation in Phonology. In A. Bruck, R. A. Fox, & M. W. La Galy (Eds.), *Papers from the Parasession on Natural Phonology* (pp. 251–274). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1992). What's cognitive, what's not, in sound change. In G. Kellermann & M.D. Morrissey (Eds.), *Diachrony within synchrony: Language history and cognition. [Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 14]* (pp. 309–355). Frankfurt am Main: Peter Lang Verlag.

O'Hara, C. (in prep.). *Soft Biases in Phonology: Learnability meets grammar.* Ph.D. dissertation. University of Southern California.

O’Hara, C. (2019). Word-Edge Consonant Database. University of Southern California.

Paradis, C. & Prunet, J.-F. (Eds.) (1991). *Phonetics and Phonology 2: The special status of coronals: internal and external evidence*. San Diego: Academic Press.

Pater, J. (1999). Austronesian nasal substitution and other NÇ̥ effects. In R. Kager, H. van der Hulst, & W. Zonneveld (Eds.), *The prosody-morphology interface* (pp. 310–343). Cambridge: Cambridge University Press.

Peperkamp, S., & Dupoux, E. (2007). Learning the mapping from surface to underlying representations in an artificial language. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 315–338). Berlin: Mouton de Gruyter.

Peperkamp, S., Skoruppa, K., & Dupoux, E. (2006). The role of phonetic naturalness in phonological rule acquisition. In D. Bamman, T.Magnitskaia, & C. Zoller (Eds.), *Papers from the 30th Boston University Conference on Language Development (BUCLD 30)* (pp. 464–475). Somerville, MA: Cascadilla Press.

Prickett, B. (2015). *Complexity and Naturalness in First Language and Second Language Phonotactic Learning*. M.A. thesis. University of North Carolina at Chapel Hill.

Prickett, B. (2018). Complexity and Naturalness Biases in Phonotactics: Hayes & White (2013) Revisited. In G. Gallagher, M. Gouskova, & S. Yin (Eds.), *Supplemental Proceedings of the 2017 Annual Meeting on Phonology*. Washington, D.C.: Linguistic Society of America.

Prince, Alan. (1990). Quantitative consequences of rhythmic organization. In M. Ziolkowski, M. Noske & K. Deaton (Eds.), *Papers from the 26th Annual Regional Meeting of the Chicago Linguistic Society, Volume 2, The Parasession on the Syllable in Phonetics & Phonology* (pp. 355–398). Chicago: Chicago Linguistic Society.

Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar.* Cambridge: Blackwell. [Published in 2004. First circulated in 1993].

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological Rule-Learning and Its Implications for a Theory of Vowel Harmony. In G. Garding & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)* (pp. 101–114). Somerville, MA: Cascadilla Press.

Rice, K. (2007). Markedness in phonology. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 79–97). Cambridge: Cambridge University Press.

Saffran, J. R. & Thiessen, E. D. (2003). Pattern Induction by Infant Language Learners. *Developmental Psychology, 39*(3), 484–494.

Schane, S. A., Tranel, B., & Lane, H. (1974). On the psychological reality of a natural rule of syllable structure. *Cognition*, *3*(4), 351–358.

Scholes, R. J. (1966). *Phonotactic Grammaticality*. The Hague: Mouton & Co.

Seidl, A. & Buckley, E. (2005). On the Learning of Arbitrary Phonological Rules. *Language Learning and Development*, *1*(3 & 4), 289–316.

Shapp, A. (2012). Substantive bias in the learning of harmony patterns. Qualifying Paper. New York University.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*(13).

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

Skoruppa, K., Lambrechts, A., & Peperkamp, S. (2011). The role of phonetic distance in the acquisition of phonological alternations. In S. Lima, K. Mullin, & B. Smith (Eds.), *Proceedings of the 39th North Eastern Linguistics Conference* (pp. 717–729). Amherst, MA: GLSA.

Smith, C. (1997). The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics*, *25*, 471–500.

Smolensky, P. (1996). The Initial State and 'Richness of the Base' in Optimality Theory. Technical Report JHU–CogSci–96–4. Johns Hopkins University.

Smolensky, P. & Legendre, G. (2006). *The Harmonic Mind: From Neural Computation to Optimality Theoretic Grammar.* Cambridge, MA: MIT Press.

Stemberger, J. P. (1983). *Speech Errors and Theoretical Phonology: A Review*. Bloomington, IN: Indiana University Linguistics Club.

Stemberger, J. P. (1991). Radical underspecification in language production. *Phonology*, *8*(1), 73–112.

Steriade, D. (1987). Redundant Values. In *Proc. of CLS 23: Parasession on Autosegmental and Metrical Phonology*, 339–362.

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.

Steriade, D. (2001b). The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. Ms. University of California, Los Angeles.

Stevens, K., Blumstein, S., Glicksman, S., Burton, M. & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America*, *91*, 2979–3000.

Sturman, B. (2018). Perceptual saliency, lenition, and learnability: An artificial grammar learning study. M.A. thesis. University of California, Los Angeles.

Terbeek, D. (1977). A cross-language multidimensional scaling study of vowel perception. *UCLA Working Papers in Phonetics, 37*.

Tessier, A.-M. (2012). Testing for OO-Faithfulness in the Acquisition of Consonant Clusters. *Language Acquisition*, *19*, 144–173.

van de Vijver, R. & Baer-Henney, D. (2014). Developing biases. *Frontiers in Psychology*, *5*(634), 1–8.

Vietti, A., Alber, B., & Vogt, B. (2018). Initial laryngeal neutralisation in Tyrolean. *Phonology*, *35*, 79–114.

Wang, M. D. & Bilger, R. C. (1973). Consonant confusions in noise: a study of perceptual features. *Journal of the Acoustical Society of America*, *54*(5), 1248–1266.

Westbury, J. & Keating, P. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*(1), 145–166.

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

Wilson, C. (2003). Experimental Investigation of Phonological Naturalness. In G. Garding & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)* (pp. 533–546). Somerville, MA: Cascadilla Press.

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

Yu, A. C. L. (2004). Explaining Final Obstruent Voicing in Lezgian: Phonetics and History. *Language*, *80*(1), 73–97.

Zhang, J., Lai, Y., & Sailor, C. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua*, *121*, 181–206.

Zhang, J., Lai, Y., & Turnbull-Sailor, C. (2006). Wug-Testing the "Tone Circle" in Taiwanese. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics (WCCFL 25)* (pp. 453–461). Somerville, MA: Cascadilla Press.