UNIVERSITY OF CALIFORNIA

Los Angeles

Bias in Phonological Learning: Evidence from Saltation

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Linguistics

by

James Clifford White

2013

ABSTRACT OF THE DISSERTATION

Bias in Phonological Learning: Evidence from Saltation

by

James Clifford White

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2013

Professor Bruce P. Hayes, Co-chair

Professor Megha Sundara, Co-chair

Understanding how people learn the phonological patterns of their language is a major challenge
facing the field of phonology. In this dissertation, I approach the issue of phonological learning
by focusing on "saltatory" alternations, which occur when two alternating sounds "leap over" an
intermediate, invariant sound (e.g., [p] becomes [v] between vowels, but [b] remains unchanged
in that context). Saltation poses a theoretical challenge because it represents excessive
modification: large perceptual changes (e.g., [p ~ v]) are licensed where small changes (e.g., [b ~
v] are not.

    I present evidence from adult artificial language experiments that saltatory systems are
dispreferred by learners. Specifically, participants who receive training data that are ambiguous
between a saltatory system and a non-saltatory system are biased towards the non-saltatory
system (Experiment 1). Moreover, when trained on a system that is explicitly saltatory,

participants find the system difficult to learn (Experiment 2). An artificial language experiment with 12-month-old infants suggests that this anti-saltation bias is also present during early language acquisition.

On the basis of the experimental results, I argue that learners have an *a priori* substantive bias that causes them to consider alternations between similar sounds to be more likely than alternations between dissimilar sounds, consistent with the principles in Steriade's (2001/2008) theory of the P-map. This bias must be a "soft" bias, rather than an absolute bias, because it must be overturned in order to learn saltations. Because saltations are attested in real languages, they must be learnable.

To account for these observations, I propose a phonological framework with three components: (1) a set of *MAP faithfulness constraints (Zuraw, 2007) that makes it possible to penalize correspondences between specific pairs of segments, (2) a substantive bias making alternations more likely if they occur between perceptually similar sounds, and (3) a Maximum Entropy learning architecture, which allows the bias to be implemented computationally via the model's prior. The proposed learning model closely matches the pattern of experimental results and it makes the right general predictions: saltations are dispreferred, but learnable given sufficient training data. More broadly, the model represents a grammatical framework that can be used to make explicit, testable predictions for future research on phonological learning. I conclude by considering the potential implications of my analysis for phonological theory, phonological acquisition, and language change.

The dissertation of James Clifford White is approved.

Robert Daland

Sharon Peperkamp

Kie Zuraw

Bruce P. Hayes, Committee Co-chair

Megha Sundara, Committee Co-chair

University of California, Los Angeles

2013

For Mom,

whose sacrifices made all of this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This dissertation simply could not have happened without the help of many wonderful people. First, I would like to offer my sincerest thanks to my advisors, Bruce Hayes and Megha Sundara. I could not have asked for a better pair of advisors to prepare me for my career, both academically and professionally. I can't count the number of times that I've walked into Megha's office to discuss something, and yet, she has always made herself available to speak with me about any topic at a moment's notice. Bruce has likewise offered invaluable assistance throughout my graduate career. He always seems to know the best way of moving forward, the right "what-ifs" to be asking, the right references to be reading, etc. I always came out of our conversations with a better understanding of my project than when the conversation started. I must say that with both Bruce and Megha, I feel like the advisor-advisee relationship has developed over time into a relationship between colleagues, and I have really enjoyed the journey.

I would also like to thank the rest of my dissertation committee. In many ways, Kie Zuraw has been like a third advisor to me and I really appreciate all of the effort that she has put in on my behalf, going all the way back to my days as an undergraduate. My experiences with both Kie and Bruce in undergrad were part of the reason that I decided to pursue linguistics (and phonology in particular) in the first place. I would also like to thank Robert Daland for having all of those informal chats with me on random thoughts (some dissertation related, some not), and also for giving me individualized help (way above the call of duty) with my programming skills. I am grateful as well for the support and advice of Sharon Peperkamp, who gladly agreed to be my outside member despite the long distance involved. Sharon's own work integrating so many

Kim, and the undergraduate research assistants who worked in the lab. Thanks also to Victoria Thatte for recording the stimuli for my infant experiment.

Special thanks go to my wonderful group of friends at UCLA, who were an invaluable part of this process: Marc Garellek, Robyn Orfitelli, Chad Vicenik, Laura Kalin, Laura McPherson, Kaeli Ward, Adam Chong, Mel Bervoets, Lauren Winans, Jianjing Kuang, Byron Ahn, Craig Sailor, and many others. You kept me going through the inevitable ups and the downs of grad school. Your support kept me cheerful and sane, and I feel blessed to have such amazing friends. Y'all are the best!

I am eternally grateful to my family, especially my mother and my sister, whose unconditional love and support have played an important role in bringing me to where I am today. You can't imagine how much it all means to me — I'm sorry if I don't let you know often enough. Part of my success is yours as much as mine.

Finally, my infinite gratitude goes to Benjamin Everly, who has been the one there to support me on a daily basis for the last 3+ years, even when I was hiding in the work cave for days on end and then grumpy for days after that. I know it wasn't always easy, but your support really did help me make it to the finish line. Thank you. Thank you. Thank you.

I'm sure that I have missed people that should be acknowledged. If you are one of them, please know that I am extremely grateful for your help and for your support.

August 22, 1986	Born, Lancaster, South Carolina, USA
2008	B.A. in Linguistics, with Departmental Highest Honors B.A. in Psychology Summa cum laude College honors University of California, Los Angeles
2008	Phi Beta Kappa
2009 – 2012	Teaching Assistant/Associate/Fellow University of California, Los Angeles
2010	UCLA Graduate Summer Research Mentorship Fellowship
2010 – 2011	UCLA Graduate Research Mentorship Fellowship
2011	M.A. in Linguistics University of California, Los Angeles
2011	UCLA Graduate Summer Research Mentorship Fellowship
2011	Lemelson Fellowship
2011 – 2012	Lenart Travel Award, UCLA
2011 – 2012	Foreign Language and Area Studies (FLAS) Fellowship U.S. Department of Education
2012	Society for Language Development Student Award
2012 – 2013	UCLA Dissertation Year Fellowship

PUBLICATIONS

Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry, 44*(1), 45–75.

Garellek, M., & White, J. (2012). Stress correlates and vowel targets in Tongan. *UCLA Working Papers in Phonetics, 110,* 65–85.

Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., & Norrmann, I. (2011). Explaining sonority projection effects. *Phonology, 28*(2), 197–234.

Garellek, M. & White, J. (2010). Acoustic correlates of stress and diagnosing syllable fusion in Tongan. *UCLA Working Papers in Phonetics, 108,* 35–65.

PRESENTATIONS

White J., & Sundara, M. Infants biases in the learning of phonological alternations. Talk presented at the 37th Boston University Conference on Language Development (BUCLD). Boston, MA.

White, J. (2012). On the learnability of saltatory phonological alternations. Poster presented at LabPhon 13. Stuttgart, Germany.

White, J. (2012). Evidence for a learning bias against "saltatory" phonological alternations in artificial language learning. Talk presented at the 86th Annual Meeting of the Linguistic Society of America. Portland, OR.

White, J., & Garellek, M. (2011). Acoustic correlates of stress and their use in diagnosing syllable fusion in Tongan. Poster presented at the 161st Meeting of the Acoustical Society of America. Seattle, WA.

White, J., & Garellek, M. (2011). Acoustic correlates of stress and their use in diagnosing syllable fusion in Tongan. Talk give at the 18th Meeting of the Austronesian Formal Linguistics Association (AFLA). Harvard University, Cambridge, MA.

White, J., & Sundara, M. (2010). Experimental evidence for phoneme-level inhibition in spoken word recognition. Poster presented at the Second Pan-American/Iberian Meeting on Acoustics/160th Meeting of the Acoustical Society of America. Cancun, Mexico.

White, J. (2009). Looking for phoneme-level inhibition in spoken word recognition using auditory lexical decision. Poster presented at the 158th Meeting of the Acoustical Society of America. San Antonio, TX.

CHAPTER 1

Introduction

A fundamental issue in phonology is determining how people acquire the phonological system of their language. As anyone who has spent a moment considering this issue (or who has just tried to learn a foreign language as an adult) will realize, the task facing the language learner is incredibly difficult. The infant's input consists of a continuous speech stream in which meaningful units such as words are rarely separated by silence. Infants must somehow track this unfriendly onslaught of signal, determine which parts of the signal are meaningful and which parts are just noise, segment discrete units (e.g., phonemes, words) from the continuous signal, and then figure out how these units are combined into the meaningful, hierarchical structures that make up language.

Making matters worse, the input contains an enormous amount of variation, some of which is systematic, and some of which is random noise. Children may hear hundreds of tokens of the "same" word, but it is unlikely that any two of those tokens are actually the same. Moreover, the input is imperfect because speakers sometimes make speech errors. Infant learners receive no overt indication about which parts of the input they should consider important and thus focus on, and which parts they should ignore. Despite these challenges, we know that children do eventually learn their native language(s); in fact, they do so at a relatively early age, and they do so (seemingly) with ease. This is the mystery of child language acquisition that excites any researcher interested in language.

1.1 Goals of the dissertation

The task facing researchers who are interested in understanding language and how it is acquired is a formidable one. For instance, we must figure out which representations are being stored by the learner but we have no direct access to those representations. We can only attempt to infer what the representations must be like based on observed outputs, that is, data collected from naturalistic settings, from corpora, or from experiments. Even gathering this type of indirect data is not straightforward. Much of linguistic knowledge is unconscious knowledge, so figuring out how to extract that knowledge is a challenge. Merely asking a speaker to be introspective, when possible, may result in data that are difficult to interpret. Any experiment, no matter how cleverly designed, comes with a range of task-specific considerations that have nothing to do with linguistic knowledge, all of which add additional noise. Moreover, the child has already acquired an enormous amount of linguistic knowledge during the first year of life (for a review, see Gervain & Mehler, 2010), before they are able to speak or complete even simple experimental tasks. Figuring out what the child knows at any given point during this period requires cleverly designed experiments and an even greater level of interpretation.

The overall point here is that language and the language acquisition process are extremely complex, and at the same time, our sources of information are indirect and noisy. In a sense, we (the researchers) are just like the child who is learning language – we have access to multitudes of data, but much of it is noise. We have to find a way to first zero in on the meaningful patterns and then figure out how those patterns should be interpreted.

The goal of this dissertation is to gain a better understanding of how people learn the phonological alternations of their language. In particular, I focus on the role that perceptual similarity plays in biasing how people learn phonological alternations. Given the complexity of

phonological learning, my approach to this problem is to focus on a specific phenomenon as a test case. My test case is a phenomenon that I will refer to as "saltation." Informally, saltations occur when an alternation between two sounds "leaps over" some other intermediate sound, such as when [p] alternates with [β], but intermediate [b] remains unchanged (saltation is defined more carefully at the beginning of Chapter 2). Saltation is a perfect test case for looking at the role of perceptual similarity in biasing phonological learning: it is attested but appears to be typologically uncommon, it cannot be derived in many mainstream phonological theories (e.g., classical Optimality Theory (OT); Prince & Smolensky, 1993/2004), and it flouts the notion that learners prefer to minimize perceptual modifications (as proposed, e.g., by Steriade, 2001/2008).

As the reader will see, this dissertation is all about saltation, and yet at the same time, it is not at all about saltation. By this, I mean that all of the empirical work (i.e., language data, experimental data, modeling) focuses on saltation; however, the saltation serves only as a vehicle for investigating larger concepts related to phonological grammars and phonological learning. As a descriptive issue alone, saltation is a phenomenon of only limited interest, relevant for only a handful of languages. But as we will see, the case of saltation has broad implications for phonological theory, both in terms of the grammatical architecture needed and how the grammar is learned.

The real strength of the approach taken in this dissertation lies in its goal of integrating pieces of evidence from several different perspectives. My goal entering this project was to understand saltation from every possible angle; some particular questions include:

- What is the status of saltation cross-linguistically?
- How does it originate and evolve historically?
- What is its status in formal phonological theory?

- How do real learners (both adults and infants) handle saltation?

- What type of phonological framework do we need to properly account for saltation as a phenomenon, both in terms of its representation in the grammar and its learnability?

As we will see, taking this approach turns out to be quite illuminating because all of the pieces fit into a coherent story. The typological and experimental data provide the facts that must be accounted for by phonological theory. The implemented learning model that I propose then provides a framework that connects the facts to the theory by making concrete, testable predictions, which can be compared to the empirical results.

In the remainder of this chapter, I will provide a brief introduction to previous research on biases in phonological learning (section 1.2) before outlining the remaining chapters of the dissertation (section 1.3).

1.2 Biases in phonological learning

There is little doubt that phonological learning is biased, in the most general sense of the word. By now, there are numerous published studies presenting cases where the outcome of phonological learning does not reflect precisely what was in the input; for instance, learners sometimes fail to pick up on statistical patterns in the input or they make certain assumptions when input is ambiguous (several of these studies are reviewed below). When the input does not match the learning outcome, we can conclude that the learning was biased in some way. This conclusion is not particularly controversial.

The controversial issue is which types of biases are at play during phonological learning and how to account for those biases in our theory of phonological acquisition. In the following sections, I will first outline the two most common experimental paradigms used to test for phonological learning biases. Then I will discuss the main types of biases that have been

proposed in the literature as well as some of the previous studies that have provided support for those biases.

1.2.1 Experimental paradigms

Experimental research testing for phonological learning biases has typically followed one of two basic paradigms: (1) a nonce-word query paradigm or (2) the artificial language learning paradigm.

1.2.1.1 Nonce-word query paradigm

In the nonce-word query paradigm, adult <u>native</u> speakers of a language are asked either to provide some type of judgment for nonce words (i.e., made-up words) in their language, or to apply some morphophonological process to nonce words. This type of test is closely related to the traditional wug-test (Berko, 1958). Typically, researchers use a corpus to find statistical patterns in the language's lexicon. In one variant of the paradigm, cases that do not appear in the lexicon are then tested to see how speakers will generalize the pattern to new types of words (e.g., Zuraw, 2007). If speakers generalize in a way that is not predicted by their language experience, we might posit an *a priori* bias that affects learning.

In another version of the paradigm, two or more patterns are chosen that crucially differ in some characteristic, such as their degree of phonetic motivation, complexity, typological support, etc., which leads the researchers to think there might be a learning bias. Speakers are then tested to see if they have internalized any (tacit) knowledge of the statistical patterns in question during the course of the native acquisition process throughout their lifetime. If speakers have failed to pick up on certain statistical patterns that hold in the lexicon, or if they have underlearned certain

patterns relative to other comparable patterns, then we can assume that their learning was biased in some way. Becker et al. (2011) referred to this particular effect as a "surfeit of the stimulus" effect because it demonstrates that the linguistic input has more richness than the learner reliably picks up on. The statistical patterns are available in the input, but the learner nevertheless fails to (fully) internalize all the patterns. Other examples of studies that have used this paradigm include Hayes et al. (2009), Becker et al. (2012), and Hayes and White (2013).

The major advantage of this paradigm is that it is possible to test for knowledge that speakers have about their native language; the results are based on linguistic competence learned through the natural language acquisition process. However, the paradigm also has limitations. A researcher is limited to investigating only patterns that are found in the languages at his or her disposal. Moreover, the researcher has no control over the linguistic input that each speaker has received over his or her lifetime and no control over other aspects of the language being tested.

1.2.1.2 Artificial language learning paradigm

The second common paradigm for testing for learning biases is the artificial language learning paradigm. In this paradigm, participants are taught an artificial (miniature) language and later tested to determine what they have learned. There are many variations on this general paradigm. Participants may be told that they are learning an unfamiliar (but real) language, a made-up language, an "alien" language, or a novel language game. The type of training may be fairly explicit, where participants are presented with sets of forms highlighting the pattern of interest (e.g., Skoruppa et al., 2011; Finley & Badecker, 2012) or it may be more implicit whereby the patterns are not made so obvious (e.g., Baer-Henney & van de Vijver, 2012). At test, participants may be provided with response options (e.g., Peperkamp & Dupoux, 2007;

6

Finley & Badecker, 2012) or participants may be required to spontaneously produce a response without access to choices (e.g., Wilson, 2006; Skoruppa et al. 2011).

Clearly, there are many possible variations on the basic paradigm; researchers choose the design specifications that fit their goals for the study. However, all such studies have the common goal of teaching participants a novel pattern in order to explore the process and/or the outcome of learning.

Beyond these details of implementation, there are two main designs for artificial language studies. In some studies, participants are merely taught two or more different patterns to see if the patterns are all (equally) learnable (e.g., Pycha et al., 2003; Peperkamp & Dupoux, 2007; Skoruppa et al., 2011; Baer-Henney & van der Vijver, 2012). A second possibility is a design that Wilson (2006) termed the "poverty of the stimulus" method: participants are trained on a pattern with critical information withheld (i.e., the training is ambiguous), and they are then tested on the cases that were withheld during training. The question in this type of experiment is not whether some patterns are more easily learned, but instead, whether learners are biased to generalize in some ways but not others. For instance, Wilson (2006) trained participants on a rule of velar palatalization either before [i] or before [e], and then tested on both vowel contexts. Participants had no way of knowing from training whether palatalization occurred in the context that they were not trained on; however, phonetic and typological considerations support generalization of palatalization from the [e] context to the [i] context, but not from the [i] context to the [e] context. To the extent that participants generalize asymmetrically in such experiments, we can conclude that their generalization was driven by some type of learning bias. Others who have used this method include Finley (2008) and Finley and Badecker (2012).

There are several advantages to the artificial language learning paradigm: the researcher can design a study to test virtually any pattern, the researcher has complete control over every aspect of the language and the learning experience, and the studies are relatively easy to run. The main disadvantage is that the language is not being acquired naturally as a first language by a child. Rather, it is being acquired by adults who already have their own native language phonologies as well as access to non-linguistic problem-solving strategies not available to young children who are learning language. Because of this, it can be difficult to know to what extent the learning reflects the same mechanisms involved in natural language acquisition. Artificial language experiments thus provide one perspective on phonological learning, but they warrant corroborating evidence from other sources. Note that it is also possible to use artificial language experiments with infants to study potential biases in early phonological learning (e.g., Saffran & Thiessen, 2003; Cristià & Seidl, 2008; Chapter 5 of this dissertation; see also Gomez & Gerken, 2000, for a review).

1.2.2 Types of biases

One particularly strong view of biased phonological learning is the theory of a universal constraint set, as assumed in classical OT (Prince & Smolenksy; 1993/2004; McCarthy & Prince, 1995) and some related theories. According to this view, humans have a universal (often considered innate) set of constraints and only phonological patterns that are derivable given some ranking of those constraints may be learned. Other patterns are deemed utterly

unlearnable.[1] The universal constraint set theory is motivated as an explanation for the typological generalizations that we find in the world's languages.

Some researchers have called on OT's universal constraint set to explain phonological bias effects found in experiments. For instance, in a nonce-word study by Becker, Ketrez, and Nevins (2011), a corpus search found that word length, place of articulation, preceding vowel height, and preceding vowel backness were all significant factors in predicting the rate of consonant laryngeal alternations in the Turkish lexicon. Turkish speakers in their experiment exhibited sensitivity to word length and place of articulation, but not features of the preceding vowel, when determining whether or not to extend the laryngeal alternations to nonce words. The authors argue that the vowel-consonant interactions observed in the corpus are not possible based on the universal set of constraints available to learners, making the vowel-consonant interactions in Turkish unlearnable (see also Becker et al., 2012). A weaker interpretation of their results would hold that learners are merely *biased* to ignore interactions between vowel features and consonant features (e.g., see Moreton, 2008), but might be capable of partially learning them (e.g., Hayes et al., 2009), or even fully learning them if such interactions were sufficiently salient in the language.

Much of the recent work looking at phonological biases has turned to exploring the existence and nature of so-called "soft" biases in phonological learning—biases that pull the learner towards certain outcomes but do not necessarily prevent the learner from acquiring a pattern. In the remainder of this section, I will discuss two biases that have been of particular interest in the literature: substantive bias and complexity bias.

---

[1] At least, as far as the grammar is concerned. Perhaps the patterns would be somewhat learnable in an experimental context if participants use extra-linguistic mechanisms to track the patterns. Moreover, in a real language, the learner might overcome this problem by simply memorizing all of the relevant word forms.

1.2.2.1 Substantive bias

One proposal is that learners tacitly take into account perceptual, articulatory, and other phonetic knowledge when learning phonological patterns.[2] Wilson (2006) termed this type of bias a "substantive bias" because it assigns phonetic substance a role in shaping how the synchronic grammar is learned. Although Wilson first introduced the term substantive bias, the concept is firmly rooted in the framework of phonetically based phonology (Hayes et al., 2004; see also Boersma, 1998; Hayes, 1999; Côté, 2000, 2004; Steriade, 1999, 2001/2008; Flemming, 2001). Substantive biases are typically conceived of as "soft" biases; the learner has an *a priori* preference for learning phonological patterns that accord well with phonetic and perceptual considerations, but the bias can be overturned in the face of contradictory data in the linguistic input.

Any phonologist would agree that phonological systems are organized in a way that strongly reflects phonetic principles. By looking at typological generalizations, it is clear that examples of this abound. For just one example, nasal assimilation (e.g., a nasal takes the place of the following consonant) occurs in language after language, presumably because nasal assimilation (a) results in less articulatory effort (one place of articulation vs. two places), and (b) has a low perceptual cost because nasal place distinctions are difficult to perceive before other consonants (Jun, 2004). On the other hand, nasal dissimilation is rare in the world's languages, presumably because it increases articulatory effort with minimal perceptual benefit.

The idea that language learners have *synchronic* biases based specifically on phonetic substance remains a controversial one. Under an alternative view, any influence of phonetics on

---

[2] This account is not necessarily incompatible with a universal constraint set. For instance, there could be a universal constraint set that is initially ranked according to phonetic principles. But there could also be phonetic biases in place in a system where the constraints are induced from the linguistic input rather than provided *a priori* (e.g., Hayes & White, 2013).

phonological systems is attributed entirely to diachrony and transmission (e.g., Ohala, 1981; Blevins, 2004, 2006; Blevins & Garrett, 2004; see also Moreton, 2008 and Moreton & Pater, 2012b, who refer to such effects as "channel bias").[3] Under this view, the nature of the human articulatory and/or perceptual system makes it such that certain sound changes are bound to occur repeatedly in language after language whereas other sound changes are highly unlikely to occur naturally. For instance, phrase-final devoicing, which occurs in many languages, might arise naturally because articulatory limitations make it difficult to maintain voicing in word-final obstruents. Moreover, perceptual limitations may lead listeners to frequently mishear final voiced obstruents as voiceless (Blevins, 2004, 2006). Thus phonological systems may diachronically evolve in ways that reflect phonetic principles even if phonetic factors do not have an active role in phonological learning at the synchronic level.

Looking only at typological generalizations is insufficient for distinguishing the evolutionary approach from the learning bias approach: both predict that phonological systems will change in ways that are consistent with phonetic principles over time. Instead, it is necessary to test learning outcomes directly, typically through experiments.

Several experiments have produced results that are consistent with a substantive bias. In an artificial language study, Wilson (2006) found that participants who were trained to palatalize velars before [e] generalized the palatalization to apply before [i] at test despite having no training in that context, but those trained to palatalize velars before [i] did not generalize to the [e] context. This asymmetrical generalization is consistent with the phonetic facts because velars

---

[3] It is widely accepted that transmission errors (e.g., mishearings, etc.) play an important role in language change. This view is not necessarily at odds with the view that inductive biases also play a role in language change (as discussed, e.g., by Moreton, 2008; Moreton & Pater, 2012b). The controversial issue is whether there are <u>substantive</u> inductive biases in addition to what Moreton calls "channel bias," or whether all phonetic influences seen in typology and language change are due to channel bias alone.

are more similar to palato-alveolars when they occur before [i] than when they occur before [e]. It is also consistent with cross-linguistic patterns of palatalization.[4] Wilson argues that the asymmetric generalization is due to a substantive bias.

Skoruppa et al. (2011) taught participants an artificial language with arbitrary phonological alternations involving a one, two, or three feature change (e.g., one feature: [p ~ t], two feature: [p ~ s], three feature: [p ~ z]). Alternations involving a single-feature change were learned more quickly and more successfully than alternations involving a two- or three-feature change (the difference in the two- and three-feature changes did not make a difference in their study). Skoruppa et al. conclude that phonetic distance affects how easy alternations are to learn, at least when it comes to sounds differing in one feature difference versus sounds differing in more features.

Hayes et al. (2009) tested speakers of Hungarian, a language with vowel harmony, on their choice of suffix allomorphs for nonce words. In some contexts, the choice of allomorph is categorically specified, but in other contexts, the choice is variable. In a corpus search, Hayes et al. found both natural constraints on the choice of allomorph (all of which were vowel harmony constraints) as well as some unnatural constraints (e.g., "prefer front suffixes when the stem ends in a sibilant").[5] When tested, the Hungarian speakers exhibited some degree of sensitivity to the unnatural constraints, but those constraints had been underlearned compared to the natural constraints.

---

[4] In fact, the interpretation of Wilson's results is somewhat problematic, as pointed out by Moreton and Pater (2012b). These issues are discussed in section 4.6.2 below.

[5] The concept of naturalness is somewhat murky and controversial in phonology. The term "natural" can be used to mean that a pattern is typologically common, phonetically motivated, or both, depending on the author and the situation. For the purposes of this review, the term "natural" can be taken to mean that a pattern is both typologically and phonetically motivated whereas "unnatural" means that a pattern has limited typological or phonetic support.

Similarly, Hayes and White (2013) found that English speakers assigned low ratings to nonce words (compared to controls) if they violated natural constraints (e.g., sonority sequencing constraints); however, unnatural constraints (e.g., no [aɪ, aʊ, ɔɪ] before [ʃ, ʒ]) had very little effect on ratings even though they have ample statistical support in the English lexicon. Because the unnatural constraints in Hayes and White's study lack phonetic motivation, the fact that they had little effect on ratings is consistent with a substantive bias. It is worth noting that Hayes and White did not find an effect of complexity in their study (see below), either defined in terms of features or natural classes, though their study was not designed to systematically test for such an effect.

Still, not every study that looks for a substantive bias effect actually finds one. Cross-linguistically, vowel harmony is very common (van der Hulst & van de Weijer, 1995), arguably due to phonetic reasons such as vowel-to-vowel coarticulation, whereas vowel disharmony is quite rare. However, a body of literature suggests that harmony and disharmony patterns are equally learnable. For instance, Pycha et al. (2003) trained participants on an artificial language with either a vowel harmony pattern (i.e., stem and suffix must agree in [back]) or a vowel disharmony pattern (i.e., stem and suffix must not agree in [back]). Both patterns were learned better than an arbitrary pattern, but there was no significant difference between the harmony and disharmony conditions. Likewise, Skoruppa and Peperkamp (2011) trained native French speakers on a novel "dialect" of French in which words underwent a process of vowel harmony (e.g., *liqueur* [likœʁ] → [likɛʁ]) or vowel disharmony (e.g., *pudeur* [pydœʁ] → [pydɛʁ]). As in Pycha et al. (2003), participants learned both the harmony pattern and the disharmony pattern better than an arbitrary pattern, but there was no difference in learnability between the harmony and the disharmony patterns. For more cases like this, see Moreton and Pater (2012b).

1.2.2.2 Steriade's P-map theory

In this dissertation, I argue that learners have a substantive bias based specifically on Steriade's theory of the P-map, short for perceptibility map (Steriade, 2001/2008). Steriade's proposal has two components. First, speakers have tacit knowledge of the relative perceptual similarity between pairs of speech sounds in any given phonological context. This knowledge is organized into a mental representation called the P-map. Second, learners are biased to prefer minimal modification – that is, they prefer phonological processes that require the smallest perceptual change. Steriade supports her claim with typological evidence, arguing that markedness violations are systematically repaired across languages by making the perceptually minimal change; for instance, in languages with bans on word-final voiced obstruents, the obstruents are overwhelmingly "repaired" by undergoing devoicing rather than by being deleted, nasalized, moved, etc. In Chapters 3 and 5, I present results from artificial language learning experiments with adults and infants, respectively, suggesting that learners really do prefer alternations between similar sounds. Other studies with similar results include Wilson (2006) and Skoruppa et al. (2011).

Steriade, working in the framework of OT, implemented the bias as an *a priori* preferred ranking of correspondence (i.e., faithfulness) constraints (see also Fleischhacker, 2005; Zuraw, 2007). However, the basic principles underlying the P-map are not applicable to only one phonological framework. In the learning model that I propose in Chapter 4, I implement a computational version of Steriade's theory, such that the learner is biased to assign greater likelihoods to alternations between perceptually similar sounds (see also Wilson, 2006, who takes a different approach to implementing the P-map computationally).

14

1.2.2.3 Complexity bias

The other type of bias proposed to affect phonological learning is a complexity bias (also known as a simplicity bias): complex patterns are more difficult to learn than simple patterns (for an extensive review, see Moreton & Pater, 2012a, 2012b). Complexity is typically judged on the basis of phonological features. Phonological patterns may be considered more complex to the extent that they (a) target classes of sounds that require more features to characterize (e.g., Pycha et al., 2003; Saffran & Thiessen, 2003; Cristià & Seidl, 2008; Skoruppa & Peperkamp, 2011), (b) require more features to change as part of a learned phonological process (e.g., Peperkamp et al., 2006b; Skoruppa et al., 2011), or (c) involve contingencies between a higher number of features (e.g., Moreton, 2008; Moreton, 2012). The proposal that complexity plays an important role in phonology goes back at least to *The Sound Pattern of English* (Chomsky & Halle, 1968), in which it was proposed that phonological rules involving fewer features are preferred to rules involving more features.

Moreton and Pater (2012b) argue that the experimental evidence in support of substantive biases in phonological learning is inconclusive at best. In many cases, complexity biases and substantive biases predict the same result, so it can be difficult to determine whether a dispreference observed for a given pattern is due to complexity or phonetic substance. For instance, consider Skoruppa et al. (2011)'s finding that alternations between sounds that differ in a single feature are easier to learn than alternations between sounds that differ in multiple features. This effect could be construed as a substantive bias based on the P-map: learners prefer minimal modification. But it could be that alternations requiring only one feature to change are less complex, and thus easier to learn, than alternations that require several features to change.

To differentiate the two biases, it is necessary to test cases in which two patterns are equal in their featural complexity, but differ from each other in terms of their phonetic motivation. The case of saltation tested in this dissertation is arguably one such case (see section 3.6.2).

1.2.3 Summary

In sum, it is clear that phonological learning is biased in certain ways. Evidence is mounting that learners have an inductive bias against complex phonological patterns, but evidence in the literature for substantive biases has been mixed (Moreton & Pater, 2012a, 2012b). Thus, an important objective for the field of phonology moving forward is to find more conclusive evidence regarding the existence of learning biases in phonology, and in particular, the status of substantive biases. In this dissertation, I explore the possibility that learners have a substantive bias to avoid non-minimal alternations, consistent with Steriade's P-map theory.

1.3 Overview of the dissertation

In **Chapter 2**, I introduce and define the phenomenon that I refer to as saltation. I provide several examples of saltations in real languages and argue that saltation must be a learnable pattern because it is attested in real languages. I present evidence that saltations arise historically from a series of independent, non-saltatory events rather than through direct saltatory sound changes, and I further conjecture that saltatory systems are unstable when they do arise in languages. Finally, I show why saltation is problematic for traditional phonological frameworks such as classical OT (Prince & Smolensky, 1993/2004), and I propose an alternative analysis making use of segment-based *MAP faithfulness constraints (Zuraw, 2007), which are themselves constrained by a bias based on the P-map.

16

In **Chapter 3**, I present results from two artificial language learning experiments with adult participants. The results indicate that saltation is a dispreferred pattern for adult learners. In Experiment 1, adults were trained on input that was ambiguous between a saltatory system and a non-saltatory system (e.g., [p ~ v], but no information about intermediate [b] or [f]). They preferred the non-saltatory system even though doing so required them to posit alternations that were not presented in the input (e.g., [b ~ v] and [f ~ v]). In Experiment 2, participants were trained on explicitly saltatory alternations (e.g., [p ~ v] with non-changing [b]). Despite their training, participants still had a tendency to change intermediate sounds in error, indicating that they found the saltatory patterns difficult to learn.

In **Chapter 4**, I propose an analysis of saltation that accounts for both its dispreferred status (based on the experimental results) and the fact that it must ultimately be learnable (because it is attested in real languages). The phonological framework consists of an implemented Maximum Entropy (MaxEnt) learning model (e.g., Goldwater & Johnson, 2003), with *MAP faithfulness constraints (Zuraw, 2007), and a P-map prior based on perceptual similarity, calculated from confusion experiments. The prior can be characterized as biasing the learner to assign greater *a priori* likelihoods to alternations between similar sounds. The model's predictions provide an excellent fit to the actual experimental results. More broadly, the model provides a framework that can make concrete, testable predictions for future studies investigating the role of perceptual similarity on phonological learning.

In **Chapter 5**, I present the results of an artificial language experiment conducted with 12-month-old infants. The results confirm that infants, like adults, disprefer saltatory alternations and instead assume that alternations are more likely to involve similar sounds. The study

17

provides evidence that the bias found in adult learners is also present during early language acquisition.

Finally, in **Chapter 6**, I conclude by summarizing the results and discussing the broader implications for phonological theory, language acquisition, and language change.

CHAPTER 2

The problem of saltation

2.1 Defining saltation

I define SALTATION as a property of phonological alternations[6], as follows:

(1) *Defn.: Saltation*

- Let A, B, and C be phonological segments.
- Suppose A and B are more phonetically similar to one another than are A and C; and B and C are also more similar to one another than A and C. In this case, B is considered INTERMEDIATE between A and C.[7]
- If in some context, A alternates with C but B remains invariant, then the alternation A ~ C is a *saltation*.

An illustrative example comes from the Campidanian dialect of Sardinian, taken from work by Bolognesi (1998). In this language, the voiceless stops /p, t, k/ are lenited to [β, ð, ɣ] when in post-vocalic position, as in (2). The voiceless affricate /tʃ/ is similarly lenited to [ʒ]. The voiced stops /b, d, g/ and voiced affricate /dʒ/ remain unchanged in that environment, as in (3). The following examples illustrate the pattern (from Bolognesi, 1998):

---

[6] In principle, there is no reason that saltation could not occur with non-alternating allophones, but all of the cases I have uncovered so far are supported by evidence from alternations.

[7] For a similar use of the term "intermediate," see Peperkamp et al., 2006a.

(2) *Post-vocalic lenition of /p, t, tʃ, k/ in Campidanian Sardinian (pp. 30–31)*

Isolation form	Post-vocalic form	Gloss
[piʃːi]	[belːu βiʃːi]	'(nice) fish'
[trintaduzu]	[sːu ðrintaduzu]	'(the) thirty-two'
[kuatːru]	[dɛ ɣuatːru]	'(of) four...'
[tʃɛʁu]	[sːu ʒɛʁu]	'(the) heaven'

(3) *Retention of post-vocalic /b, d, dʒ, g/ (pp. 36–39)*

Isolation form	Post-vocalic form	Gloss
[bĩu]	[sːu bĩu]	'(the) wine'
[gɔma]	[dɛ gɔma]	'(of) rubber'
[dominiɣu]	[donːja dominiɣu]	'(every) Sunday'
[dʒikːoɾja]	[dɛ dʒikːoɾja]	'(of) chicory'

Bolognesi attests to the productivity of the pattern with examples of application to borrowed

or recently introduced words:

(4) *Productivity of the alternation in borrowed words*

Isolation form	Post-vocalic form	Gloss
[polonia]	[sːa βolonia]	'(the) Poland'
[tasːi]	[sːu ðasːi]	'(the) taxi'
[komputːɛ]	[sːu ɣomputːɛ]	'(the) computer'

An apparent exception to this productivity is the affricate [tʃ], which avoids lenition in Italian

loans (e.g., [sːu tʃinɛma] 'the cinema', *[sːu ʒinɛma]). Bolognesi further notes (p. 36) that the

output pattern is maintained consistently: "Speakers not only do not spirantize voiced stops, but

judge this ... as entirely ungrammatical, instead. For them a phrase such as, for example, *sa:*

βɔt:a could only be the output of underlying *sa: pɔrta* ('the door'), and never of *s:a bɔrta* ('the time'). They claim the second interpretation to be wrong."

The term "saltation" is borrowed from Minkova (1993) and Lass (1997), who use the term to describe similar cases in the context of historical sound change.[8] Derived from the Latin word for "leaping," the term is apt in this case because [p] intuitively leaps over intermediate [b] in order to reach [β]; sounds at the other places of articulation in (2) and (3) behave comparably. Figure 1 illustrates the "leaping" behavior involved in such alternations.

Figure 1. Example of a saltation, as in Campidanian Sardinian.
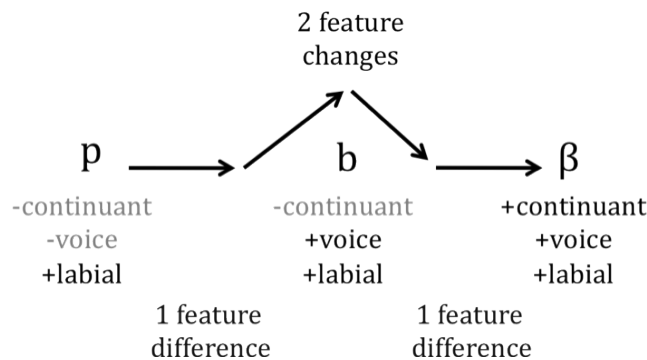
Returning to the definition in (1), the classification of the [p ~ β] alternation in Campidanian Sardinian (and the other alternations in (2) as well) as saltatory is based on the following observations:

---

[8] This phenomenon has been discussed before by Lubowicz (2002) and Ito and Mester (2003) as "derived environment effects," a term that is intrinsically tied to their proposed analysis of the problem (see section 2.5.1 below). The term "saltation" is used here in a purely descriptive way; it refers to the phonological pattern itself without referencing any particular analysis.

- [p] differs from [β] in two phonological features (voicing and continuancy).

- [b] differs from [p] in only one of those features (voicing) and [b] differs from [β] in only one of those features (continuancy); thus, [b] is more similar to [p] and more similar to [β] than [p] and [β] are to each other. Therefore, [b] is intermediate between [p] and [β].

- [p] alternates with [β], whereas intermediate [b] is invariant. This is a saltation.

Note that the definition of saltation given in (1) requires that the similarity between pairs of sounds be judged according to some phonetic measure. Perhaps the most straightforward measure is to use phonological features, as I have done with the Campidanian Sardinian case in the previous discussion. In the case of phonological features, the definition in (1) can be restated in terms of a subset relationship with respect to features: a sound B is intermediate between sounds A and C if the set of features for which A and B differ, and the set of features in which B and C differ, are both subsets of the set of features for which A and C differ. The definition given in (1) is more general because any phonetic continuum may be used. In Chapter 4, I will argue that perceptual similarity is the appropriate continuum for this purpose, so the more general definition is warranted.

2.2 Cases of saltation in the world's languages

To my knowledge, there has been no systematic typological study of saltation. The following sections provide a brief overview of the cases of saltation that I am aware of in the world's languages (in addition to the case in Campidanian Sardinian, described above).

2.2.1 German (Ito and Mester, 2003)

In German, coda /g/ following atonic [ɪ] surfaces as [ç], which is the allophonic variant of [x] that appears after front vowels:

(5) *Underlying coda /g/ surfaces as [ç] after atonic [ɪ]*

    a. /kø:nɪg/   →   [kø:nɪç] 'king'   (cf. [kø:nɪgə] 'kings')
    b. /ho:nɪg/   →   [ho:nɪç] 'honey'   (cf. [ho:nigə] 'honey', dat.)
    c. /ve:nɪg/   →   [ve:nɪç] 'little'   (cf. [ve:nɪgə] 'few')

According to Ito and Mester, the same process of spirantization applies to every coda /g/, not just the ones after [ɪ], in Colloquial Northern German, a "regional standard" variety spoken in northern Germany. The resulting fricative is either [x] or [ç] depending on the frontness of the preceding vowel:

(6) *Every underlying coda /g/ surfaces as [x] or [ç] in Colloquial Northern German*

    a. /tʀu:g/   →   [tʀu:x] 'carried', 1sg (cf. [tʀu:gən] 'carried', 1pl)
    b. /fʀa:g/   →   [fʀa:x] 'asked', 1sg   (cf. [fʀa:gən] 'asked', 1pl)
    c. /ve:g/   →   [ve:ç] 'way'   (cf. [ve:gə] 'ways')

In both cases, underlying coda /k/ surfaces as [k] without undergoing the spirantization:

(7) *Underlying /k/ remains [k]*

    a. /plastɪk/ →   [plastɪk], *[plastɪç]   'plastic'
    b. /ba:tɪk/   →   [ba:tɪk], *[ba:tɪç]   'batik'

Note that [k] differs from [g] in voicing and from [x] in continuancy, whereas [g] and [x] differ in voicing and continuancy. Thus according to the definition in (1), the alternation is an example of saltation because non-alternating [k] is intermediate between alternating [g ~ x] (and likewise between [g ~ ç]).

2.2.2 Polish (Lubowicz 2002)

In Polish, underlying velars are realized as post-alveolar affricates before front vowels.

However, underlying /g/ is also spirantized in that context, surfacing as [ʑ] instead of [ʥ].

Underlying /ʥ/, however, remains unaffected; it does not undergo spirantization before front

vowels (Lubowicz, 2002; Rubach, 1984):

(8) a. *Underlying /g/ surfaces as [ʑ] before front vowels*

| /va[g]+i+ć/ | → | [va[ʑ]+ɨ+ć] | 'to weigh' |
| /śńe[g]+ĭc+a/ | → | [śńe[ʑ]+ɨc+a] | 'snowstorm' |

b. *Underlying /ʥ/ remains [ʥ]*

| /brɨ[ʥ]+ĭk+ĭ/ | → | [brɨ[ʥ]+ek], * [brɨ[ʑ]+ek] | 'bridge (dim.)' |
| /[ʥ]em+ĭ/ | → | [[ʥ]em], *[[ʑ]em] | 'jam' |

In this case, the affricate [ʥ] differs from [g] in place of articulation and stridency and it differs

from [ʑ] in continuancy. Alternating [g] and [ʑ] differ in all three of those features, so the

alternation saltates over intermediate [ʥ].

As Rubach notes (p. 121), the forms with unchanging /ʥ/ are foreign borrowings into the

language. This point is discussed further in section 2.3.

2.2.3 *Manga dialect of Kanuri (Hutchinson, 1981; Jarrett, 2007)*

In the Manga dialect of Kanuri, underlying /t/ surfaces as [ð] when it occurs after a

[+sonorant] sound (i.e, vowels or sonorant consonants) and before a vowel. Alternations occur

with the *–tú* suffix, which is reported to be the class 2 verbal infinitive marker by Hutchinson

(1981):

(9) a. */t/ surfaces as [ð] after vowels or sonorant consonants*

/lá + tú/	→	[láðú]	'to dig, shovel'
/kà + tú/	→	[kàðú]	'to avoid, escape'
/fàn + tú/	→	[fànðú]	'to hear, feel'
/kàl + tú/	→	[kàlðú]	'to change, exchange'
/fér + tú/	→	[férðú]	'to spread, lay out'

  b. */t/ surfaces as [t] elsewhere*

/dàp + tú/	→	[dàptú]	'to refuse, prohibit'
/də́p + tú/	→	[də́ptú]	'to divorce'
/kòk + tú/	→	[kòktú]	'to peck (e.g., a hen)'
/táp + tú/	→	[táptú]	'to fill (with liquid)'
/tùs + tú/	→	[tùstú]	'to rest, remain'

The nominalizing prefix *kə̀n* (Hutchinson, 1981, pp. 77–78) also triggers spirantization of /t/ to

[ð] (10a), but underlying /d/ does not undergo spirantization (10b):

(10)   a. *Spirantization of underlying /t/*

        [tà]   'seize, catch (verb)'   →   [kə̀nðâ]   'catch (noun)'

    b. *No spirantization of underlying /d/*

[dóndì]	'sick, ill'	→	[kə̀ndóndì]	'sickness'
[də̀gà]	'live, stay (verb)'	→	[kə̀ndə́gà]	'way of life'
			(also: [kə̀ndə̀gàrám] 'dwelling place')	

This case is similar to the case observed in Campidanian Sardinian (section 2.1), where the same type of lenition occurs at every place of articulation. It is saltatory because [t] alternates with [ð], jumping over intermediate non-alternating [d].

2.2.4 Rendaku in Tokyo Japanese (Ito & Mester, 1997)

According to Ito and Mester (1997), the conservative Tokyo dialect of Japanese has a pattern of *rendaku* that could be considered optionally saltatory. Generally speaking, *rendaku* is a process by which a voiceless obstruent becomes voiced if it occurs at the beginning of the second word in a compound (e.g., [tama] 'ball' → [teppo:dama] 'gun ball'). In the conservative Tokyo dialect discussed by Ito and Mester, voiced velars also undergo a process of nasalization when they are not at the beginning of a prosodic word.

The processes of *rendaku* and velar nasalization interact in an interesting way. When underlying /k/ appears as the initial segment in the second stem of a compound, it obligatorily undergoes velar nasalization; thus /k/ → [ŋ]. However, when underlying /g/ appears as the first segment of the second stem in a compound, it may be <u>optionally</u> nasalized to [ŋ] or it may remain [g]. This results in possible minimal pairs, such as the ones reported by Ito and Mester in (11a) and (11b):

(11) a. *Underlying /k/ surfaces as [ŋ]*

| /kama/ | 'kettle' | → | [o:ŋama] | 'big kettle' |
| /kumi/ | 'class' | → | [kiŋumi] | 'yellow group' |

b. *Underlying /g/ optionally remains [g]*

| /gama/ | 'toad' | → | [oːgama] ~ [oːŋama] | 'big toad' |
| /gumi/ | 'berry' | → | [kigumi] ~ [kiŋumi] | 'yellow berry' |

[k] and [ŋ] differ in voicing and nasality, whereas intermediate [g] differs from [k] in voicing and from [ŋ] in nasality. Thus, in the cases in which /g/ remains optionally unchanged, the [k ~ ŋ] alternation is saltatory.

2.2.5 *Russian vowel reduction (Crosswhite, 2000)*

Russian has extensive vowel reduction in unstressed syllables. In dialects spoken in central Russia, including the dialect known as Contemporary Standard Russian, unstressed /o/ reduces to [a], and unstressed /e/ reduces to [i], in immediately pretonic syllables (unstressed /o/ and /a/ further reduce to [ə] in non-pretonic position). Following palatalized consonants, /e, a, o/ all reduce to [i] (Crosswhite, 2000). This pattern is illustrated in Figure 2.

Figure 2. Vowel reduction pattern for pretonic position in Contemporary Standard Russian (Crosswhite, 2000).
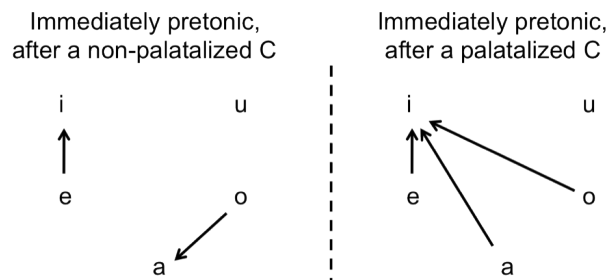
Of interest here is that unstressed /o/ changes to [i] after palatalized consonants, but unstressed /u/ does not change in the same context. [o] differs from [i] in height, rounding, and backness whereas [u] differs from [i] in only a subset of those features – rounding and backness

27

but not height. Therefore, the /o/ → [i] change is a saltation, jumping over /u/. Indeed, Crosswhite (2000) points out that her basic OT analysis predicts that /u/ will also reduce to [i]; she proposes allowing conjoined constraints (see section 2.5.1), or considering the [o ~ i] alternation to be lexicalized, as possible solutions.

2.2.6 Slovak diphthongization

Lubowicz (2002, p. 249), relying on Rubach (1993), suggests that the length alternations involving diphthongization in Slovak are saltatory. Specifically, various processes of lengthening and shortening result in short [e] and [o] alternating respectively with [ie] and [uo]. Slovak also has phonemic [eː] and [oː], which remain invariant in quality. Depending on one's analysis (including Rubach's), this phenomenon could be saltatory. Assuming that [eː], [oː], [ie], and [uo] all share a feature [+long] and that [e], [o], [eː], and [oː] all share a feature marking them as monophthongs (to the exclusion of [ie] and [uo]), [eː] and [oː] would be intermediate between alternating [e ~ ie] and [o ~ uo], respectively.

2.2.7 Suma (Bradshaw 1995, 1998)

In the associative construction, a morphosyntactic construction in which two sequential nouns are related to each other, a final low tone becomes high when preceded by a high tone, resulting in an alternation between a HL pattern and a HH pattern:

(12) *HL ~ HH tonal alternation in Suma*

kpánà	'jar'	→	kpáná rì	'water jar'
ɓólò	'pouch'	→	ɓóló náŋá	'animal skin pouch'
kúrì	'egg'	→	kúrí gɔ̌k	'serpent's egg'

Bradshaw (1998) states that "nouns with final H or M tones do not alternate" (p. 117); however, no examples of this type are given. If indeed the HL and HH patterns alternate while the HM pattern does not alternate, the L~H tonal alternation in the Suma associative construction can be considered saltatory: a low tone must jump over a mid tone to reach a high tone.[9]

2.2.8 Summary

Table 1 provides a summary of the saltations discussed above. This list of saltations is almost certainly incomplete, but includes the cases of which I am currently aware. Before moving on, it is worth briefly considering the overall impression made by these cases.

The "best" case of saltation in Table 1 is likely the one from Campidanian Sardinian because it has three characteristics: (a) it applies to a general class of sounds (i.e., all voiceless stops and affricates), (b) it is obligatory, and (c) it is reported by Bolognesi to apply productively to words that have been recently introduced into the language. By contrast, the other cases in Table 1 lack one or more of those characteristics. Most of the other languages have saltatory alternations involving only a single phoneme as the target, rather than an entire class of sounds (with the exception of Slovak). Moreover, the tonal case in Suma is limited to a very specific morphological construction, and the intermediate [g] in the Japanese *rendaku* case only optionally remains unaltered. In addition, some cases appear to have limited or no productivity

---

[9] Kie Zuraw (personal communication) points out that saltations might occur frequently in patterns of tone sandhi. To address this possibility, we would first need to determine precisely how to characterize saltation in the tonal domain, which is not trivial. For instance, does a low level tone that changes to a high falling tone saltate over a mid rising tone? What if there are secondary voice quality cues involved? Much of this will depend on the feature system assumed. It also remains unclear how human learners respond to (potential) saltation in the tonal domain. These questions warrant investigation in future research.

for new words entering the language (e.g., the Russian vowel reduction case, see section 2.3 below).

Overall, the evidence seems to suggest that saltation is a rare phenomenon in the world's languages. All of the known cases have apparent quirks, raising questions about how saltation comes to exist in a language as well as about its status once it does exist. I return to this issue in the next section. Based on evidence presented in the remainder of the dissertation, I conjecture that saltation is a dispreferred pattern for the language learner, and I predict that saltatory systems, when introduced into a language, are unstable over time.

Table 1. Summary of cases of saltation. Saltation column lists the saltatory alternation(s) in the language with the intermediate sound being saltated over in parentheses. Language family information comes from Ethnologue (Lewis et al., 2013).

Language	Language family	Type of (supra-) segments involved	Saltation(s)	Reference(s)
Campidanian Sardinian	Indo-European (Italic)	consonants	p ~ β (b) t ~ ð (d) tʃ ~ ʒ (dʒ) k ~ ɣ (g)	Bolognesi 1998
German	Indo-European (Germanic)	consonants	g ~ ç (k)	Ito & Mester 2003
Polish	Indo-European (Slavic)	consonants	g ~ ʑ (ɖʐ)	Lubowicz 2002
Kanuri (Manga dialect)	Nilo-Saharan	consonants	t ~ ð (d)	Jarrett 2007, Hutchinson 1981
Tokyo Japanese	Japonic	consonants	k ~ ŋ (g – optionally)	Ito & Mester 1997
Russian	Indo-European (Slavic)	vowels	o ~ i (u)	Crosswhite 2000
Slovak	Indo-European (Slavic)	vowels	e ~ ie (eː) o ~ uo (oː)	Lubowicz 2002
Suma	Niger-Congo	tones	HL ~ HH (HM)	Bradshaw 1995, 1998

2.3 Historical development of saltatory phonological systems

Minkova (1991) and Lass (1997) claim that diachronic sound changes are never saltatory. Lass goes further, saying that sound changes never leap over even hypothetical segments, much less segments that actually exist in a language. But how is it that saltations come to exist in the first place? Hayes and White (in prep.) argue that saltatory systems are never innovated directly; instead, they arise by accident, through a series of non-saltatory sound changes, grammatical restructuring, and introduction of new vocabulary. Here, I will provide a brief sketch of the evidence for this claim, but I refer the reader to the paper by Hayes and White for more details.

The proposal is that saltations come into existence from one of three different historical mechanisms. In one possible case, intermediate sounds are reintroduced due to grammatical restructuring. For instance, the alternation [p ~ β] in Campidanian Sardinian seems to have resulted from two parallel, diachronic lenition chains, one affecting underlying voiceless stops after vowels, p → b → β, and one affecting voiced stops, b → β → ∅ (and comparably for the other places of articulation – I use only labials here for expositional purposes). At the endpoint, the system was such that /p/ → [β] and /b/ → ∅, with non-saltatory [p ~ β] alternations, but extreme neutralization of all voiced consonants. At a later point, [b] was reintroduced in the post-vowel position due to grammatical restructuring, resulting in a system with saltatory [p ~ β] alternations. There is evidence for the pre-saltatory historical stage in modern Campidanian Sardinian. Bolognesi reports (p. 37) that some *b*-initial words in modern Campidanian Sardinian still alternate with zero in free variation, though the allomorph with initial [b] intact is more common than the allomorph without [b] in spontaneous fluent speech. The *b*-initial words that allow alternation with zero appear to be common, everyday words, making it plausible that the

31

allomorphs without initial [b] are memorized forms representing remnants from an older stage of the language.

The second proposed origin of saltation is when intermediate sounds are newly introduced through foreign borrowings. For example, Rubach (1984, p. 121) points out that the cases of intermediate /ʥ/ before front vowels in Polish are in borrowed words such as [briʥek] 'bridge (dim.)' and [ʥem] 'jam.' Similarly, Rubach (1993, p. 177) points out that intermediate [eː] and [oː] entered Slovak as cosmopolitan loanwords, as in [treːn] 'military carriage' and [moːda] 'fashion.' Cases of [ɪk] in German also originated as loanwords, e.g. in [plastik] 'plastic' and [baːtɪk] 'batik.'

Finally, saltations can be created when two sounds undergo separate sound changes, but end up flanking an already existing intermediate sound as a result. As reported by Crosswhite (2000, p. 167) this apparently happened in the case of the saltatory [o ~ i] alternation in Russian. Historically, stressed [e] changed to [o] when preceded by a palatalized consonant and not followed by one. Unstressed [e] remained [e] in that context, resulting in alternations between stressed [o] and unstressed [e]. Through an independent sound change, unstressed [e] was eventually raised to [i] after palatalized consonants. The result was a system in which stressed [o] alternated with unstressed [i], saltating over intermediate [u].

As a preview, I further conjecture that saltation is an unstable pattern when it appears in a language. In Chapter 3, I present experimental evidence that saltation is a dispreferred pattern for language learners. There is also real language evidence that such patterns are unstable. Crosswhite (2000, p. 168) reports that the saltatory [o ~ i] alternation is losing its productivity in modern Russian. By conducting an informal experiment with native speakers, she concludes that

speakers are reluctant to extend the alternation to new words. She argues that the alternation is moving towards becoming a memorized morphological pattern rather than a productive phonological process. Though we have little data investigating the diachronic stability of saltatory systems, I imagine that we would find more cases like this if such investigations were undertaken.

2.4 The problem of deriving saltation in phonological theory

I turn now to the question of how to generate saltations in phonological theory. It is clear that any workable phonological theory must be able to generate saltations because they are attested in natural languages (see section 2.2). In this section, I demonstrate why saltation cannot be derived in classical Optimality Theory (OT; Prince & Smolensky, 1993/2004), as well as in two closely related theories, Harmonic Serialism (McCarthy, 2000) and Harmonic Grammar (Legendre, Miyata, & Smolensky, 1990; Smolensky & Legendre, 2006; Pater, 2009b). Ultimately, we will see that the real problem with deriving saltation is a problem with the constraint set assumed in classical OT.

2.4.1 Classical Optimality Theory

By classical OT, I mean OT as implemented by Prince & Smolensky (1993/2004), augmented by the Correspondence Theory of McCarthy and Prince (1995). Classical OT cannot generate this type of saltatory process, as has been shown elsewhere by Lubowicz (2002) and Ito and Mester (2003), who referred to the same phenomenon as "derived environment effects" (see section 2.5.1 below).

The problem, in essence, arises from the excessive nature of saltation – a large change happens where a small change does not. Consider the case in Campidanian Sardinian, where /p, t, k/ becomes [β, ð ɣ] between vowels. In order for /p/ to become [β], /p/ must undergo two changes: voicing and spirantization. To get each of these changes, both of the relevant markedness constraints must be ranked above the relevant faithfulness constraints, giving the necessary constraint ranking in (13):

(13)   *V[−voice]V, *V[−continuant]V   >>   IDENT(voice), IDENT(continuant)

That is, it is more important to avoid having voiceless sounds or stops between vowels than to change either of those features. As shown in the tableau in (14), this ranking generates the proper output, [VβV] from underlying /VpV/. The faithful candidate, [VpV], as well as both of the intermediate candidates, [VbV] and [VɸV], are ruled out by one or both of the highly ranked markedness constraints:

(14)

/VpV/	*V[−voice]V	*V[−cont]V	IDENT(voice)	IDENT(cont)
☞ VβV			*	*
VbV		*!	*	
VɸV	*!			*
VpV	*!	*		

However, a problem arises when applying the same constraint ranking to the input /VbV/, which should surface unaltered as [VbV]:

34

(15)

/VbV/	*V[−voice]V	*V[−cont]V	IDENT(voice)	IDENT(cont)
☞ *VβV				*
☹ VbV		*!		
VφV	*!		*	*
VpV	*!	*	*	

As tableau (15) shows, the candidate [VβV] is still favored because it satisfies both of the high-ranked markedness constraints. To get the correct candidate [VbV] to win, the faithfulness constraint protecting the [b] from spirantizing, IDENT(cont), would need to be ranked above *V[−cont]V. Doing so, however, would also incorrectly cause [VbV] to win when the input is /VpV/. Therefore, we have a ranking paradox – a different constraint ranking is required in order to get the proper outputs for each of the two inputs, that is, for /p/ to surface as [β], but intermediate /b/ to surface unaltered as [b].

More generally, this situation would occur in any case of saltation, as defined in section 2.1. Consider any generic case where A becomes C, jumping over intermediate, unchanging B. The faithfulness violations caused by B → C are necessarily a subset of the violations caused by A → C, so if A → C is allowed, then B → C must also occur. It is impossible to promote the faithfulness constraint that would protect B over the relevant markedness constraint because doing so would result in A → B rather than A → C. In sum, there is no way to derive saltation in classical OT.

2.4.2 Harmonic Serialism

Harmonic Serialism (HS; McCarthy, 2000, *et seq.*) is a derivative of classical OT in which intermediate outputs are evaluated at multiple steps before a final output is reached. Candidates are evaluated according to strictly ranked constraints, as in classical OT; however, in any given

35

step, only candidates involving a single change are considered (i.e., changes at each step <u>must</u> be minimal). The output from each step becomes the input to the immediately following step, and constraint rankings remain constant between steps. This sequential process continues until the model converges at the cycle in which the best candidate is fully faithful to the input of that particular step (i.e., no further changes are made). This final winning candidate is the overall output. The result of this process is a sort of phonological path through candidates, with each candidate representing an incremental progression from the original input form to the final output form.

   HS, restricted to the constraint set assumed in classical OT, is unable to generate saltations, essentially for the same reason that classical OT cannot. The reasoning becomes perhaps even more apparent in HS because of the notion of a phonological path. Consider the test case from Campidanian Sardinian: /p/ → [β], /b/ remains [b]. In order to get unchanging /b/, IDENT(cont) must outrank *V[−cont]V, as shown in (16):

(16)   *Input /b/ surfaces unaltered as [b]*

      Step 1 (convergence):

/VbV/	IDENT(cont)	*V[−cont]V	*V[−voice]V	IDENT(voice)
☞ VbV		*		
VβV	*!			
VpV		*	*!	*

       Final output: [VbV]

36

But under this ranking, intermediate [b] quite literally <u>blocks</u> the path from /p/ to [β]. By ranking IDENT(cont) high to protect /b/, there is no way for /p/ to reach [β] because the derivation gets "stuck" at [b]:

(17) *Underlying /p/ gets "stuck" at [b] in Step 2*

Step 1:

/VpV/	IDENT(cont)	*V[–cont]V	*V[–voice]V	IDENT(voice)
☞ VbV		*		*
VpV		*	*!	

Step 2 (convergence):

/VbV/	IDENT(cont)	*V[–cont]V	*V[–voice]V	IDENT(voice)
☞ VbV		*		
☹ VβV	*!			
VpV		*	*!	*

Final output: *[VbV]

In order to have /p/ surface as [β] in HS, there must be an intermediate stage in which [b] changes to [β]. However, this doing so will result in the same outcome when the input to <u>any</u> stage is /b/, regardless of whether the original input is /p/ or /b/; indeed, note that Step 2 in (17) is (and must be) identical to Step 1 in (16). Thus, it is impossible to have /p/ → [β], but /b/ → [b].

One might think it possible to overcome the problem in HS by taking a different "path." There are two possible paths leading from /p/ to [β] depending on which sound serves as the output of the first step: the "b-path" and the "ɸ-path." If the /p/ → [β] change in Campidanian Sardinian could go through the ɸ-path, it might be possible to avoid the intermediate [b] altogether, leaving underlying /b/ unaltered. Moreover, this solution is tempting because there

37

is no [ɸ] in Campidanian Sardinian; thus we do not have to worry about protecting intermediate

/ɸ/ from changing in the same way that we must protect intermediate /b/ from changing. A

closer look, however, reveals that this possibility does not solve the saltation problem in HS.

As mentioned above, in order for /b/ to remain [b] rather than changing to [β], IDENT(cont)

must outrank *V[–cont]V:

(18) *Underlying /b/ surfaces as [b]*

Step 1 (convergence)

/VbV/	IDENT(cont)	*V[–cont]V
☞ VbV		*
VβV	*!	

Final output: [VbV]

With the input /VpV/, however, the same ranking also protects underlying /p/ from changing to

[ɸ]. Even if we prevent /p/ from going down the b-path by ranking IDENT(voice) high, the result

is that candidate [p] is preferred over candidate [ɸ] as the output of the first step:

(19) *IDENT(cont) >> *V[–cont]V also protects underlying /p/*

Step 1 (convergence):

/VpV/	IDENT(voice)	*V[–voice]V	IDENT(cont)	*V[–cont]V
VɸV		*	*!	
VbV	*!			*
☞ VpV		*		*

Final output: *[VpV]

38

Even if we force a change by ranking a constraint against intervocalic voiceless stops (i.e.,

*V[–voice, –cont]V) at the top of the grammar, it is still not possible to get through the ɸ-path.

In Step 2, any constraint favoring [ɸ] must be ranked below constraints favoring [β]. In

particular, IDENT(voice), which favors [ɸ], must be ranked below *V[–voice]V, which favors [β].

The exact opposite ranking, however, is necessary to avoid that b-path in Step 1. We have a

ranking paradox:

(20) *A ranking paradox*

a. IDENT(voice) >> *V[–voice]V: needed for p → ɸ, not p → b
b. *V[–voice]V >> IDENT(voice): needed for ɸ → β, not ɸ → ɸ

In sum, there is no way to derive saltation in HS using the classical set of constraints.

2.4.3 Harmonic Grammar

Another approach that we might take is to abandon the strict ranking requirements of

classical OT in favor of a system with weighted, additive constraint evaluation, as in Harmonic

Grammar (HG; e.g., Legendre, Miyata, & Smolensky, 1990; Smolensky & Legendre, 2006;

Pater, 2009b). In HG, each constraint has an associated weight instead of a ranking. At the time

of evaluation, each violation is first multiplied by the weight of the constraint violated. The

resulting figures are summed for each candidate, resulting in the candidate's harmony score. The

harmony score is often made negative, reflecting the fact that it is actually a penalty, and the

candidate with the harmony score closest to 0 (i.e., the smallest penalty) is considered the

winner.

Appealing to the weighted constraints in HG is appealing because such grammars have the property of "ganging"; that is, multiple violations of lowly ranked constraints may team up to overcome a single violation of a constraint with a higher weight. This outcome is never possible with strict ranking as in classical OT. Indeed, it has been shown that ganging can account for some (but not all) of the effects that have previously been analyzed with conjoined constraints (Legendre et al., 2006; Pater, 2009a, 2009b; Potts et al., 2009), and conjoined constraints are capable of deriving saltation (see section 2.5.1 below).

A closer look reveals that HG is also unable to account for saltation. Consider the tableaux in (21) below (columns for constraints that do not differentiate the two crucial candidates have been shaded). In order for underlying /b/ to surface unaltered as [b] instead of [β], IDENT(cont) must have a higher weight than *V[−cont]V because those are the only two constraints that differentiate the candidates (21a). When the input is /p/ as in (21b), the two crucial candidates are also [b] and [β] (the faithful candidate [p] is ruled out by highly weighted *V[−voice]V). In this case, both candidates violate IDENT(voice), so it cannot affect the outcome no matter what its weight is set to; likewise, neither candidate violates *V[−voice]V so it cannot affect the outcome. Thus, only IDENT(cont) and *V[−cont]V can differentiate the two candidates. The problem is that there is no way to weight the constraints in order to get the correct outcomes in both tableaux. IDENT(cont) needs a higher weight to protect underlying /b/, but *V[−cont]V needs a higher weight to force underlying /p/ to change to [β]. Even with weighted constraints, we see that the problem boils down to the same issue as in classical OT: there is no way to rank and/or weight the constraints to get the desired outcomes.

(21) *Harmonic Grammar is unable to derive saltation.*

a. IDENT(cont) *must have a higher weight than *V[–cont]V for /b/ to remain unaltered.*

/VbV/	*V[−voice]V	IDENT(cont)	*V[−cont]V	IDENT(voice)	Harmony
	4	3	2	2	
☞ VbV			−1		−2
VβV		−1			−3
VpV	−1		−1	−1	−8

b. **V[–cont]V must have a higher weight than* IDENT(cont) *for /p/ → [β].*

/VpV/	*V[−voice]V	IDENT(cont)	*V[−cont]V	IDENT(voice)	Harmony
	4	3	2	2	
☞ VbV			−1	−1	−4
☹ VβV		−1		−1	−5
VpV	−1		−1		−6

2.5 Modifying the set of constraints to account for saltation

The problem of deriving saltation is not so much a problem for the architecture of OT, but rather a problem with the set of constraints assumed under classical OT (McCarthy & Prince, 1995). To have saltation, it must be possible for a long journey to be licensed where short journeys are not. In this section, I consider two possible ways of updating the constraint set so saltation is allowed in OT. The first proposal, offered by Lubowicz (2002) and Ito & Mester (2003), is to allow local constraint conjunction. Though it makes saltation possible, I argue that constraint conjunction is not the theory that we want. I then present (and ultimately adopt) an analysis using *MAP constraints (Zuraw, 2007), which are themselves constrained by the theory of the P-map (Steriade, 2001/2008).

2.5.1 The local constraint conjunction solution

Lubowicz (2002) and Ito and Mester (2003) solve the problem of saltation by appealing to local constraint conjunction. The idea behind local constraint conjunction is that two constraints can be combined to form a single constraint, which is violated only when both sub-constraints are violated in the same domain (Smolensky, 1993, 1995, 2006). As demonstrated by Lubowicz and Ito and Mester, saltation can be derived in OT so long as local constraint conjunction is allowed to combine a markedness constraint and a faithfulness constraint. In effect, it allows a markedness constraint to penalize a marked segment that is not faithful to the underlying form while overlooking the same marked segment when it is present underlyingly. To derive the saltation in Campidanian Sardinian, we must combine the markedness constraint *V[−cont]V and the faithfulness constraint IDENT(voice) into the conjoined constraint [*V[−cont]V & IDENT(voice)]Seg. The conjoined constraint is violated whenever a [−cont] segment that occurs between vowels is also unfaithful in its voicing feature.

By ranking the conjoined constraint high in the grammar, it is possible to derive saltation, as shown in (22):

(22)   a. */p/ correctly surfaces as [β] due to the conjoined constraint*

/VpV/	[*V[−cont]V & ID(voice)]Seg	*V[−voice]V	ID(voice)	ID(cont)	*V[−cont]V
☞ VβV			*	*	
VpV		*!			*
VbV	*!		*		*

b. *Underlying /b/ does not violate the conjoined constraint and is protected*

/VbV/	[*V[−cont]V & ID(voice)]Seg	*V[−voice]V	ID(voice)	ID(cont)	*V[−cont]V
VβV				*!	
☞ VbV					*

42

For input /VpV/, the high ranking conjoined constraint rules out candidate [VbV] because [b] is a

stop that has also changed its voicing (22a). This allows IDENT(cont) to be ranked above

*V[–cont]V, which protects underlying /b/ from changing into [β] (22b). Candidate [VbV] does

not violate the conjoined constraint when /VbV/ is underlying because it is faithful in voicing.

The intuition behind these conjoined constraints, as used by Lubowicz, is that a segment

should not be marked if it is also unfaithful, or put a different way: if the segment is going to

change anyway, it might as well change into something even better (i.e., something less marked).

This property leads Lubowicz to refer to saltations as "derived environment effects" (related to

the "derived environment rules" of Kiparsky (1973)) because under her analysis, the

modification (in this case, b → β) only affects segments that have already been derived from

something else.

The intuition behind Lubowicz's analysis is reasonable, but the proposed formalism has

major, potentially dire, consequences for phonological theory in general. As Hayes and White (in

prep.) argue, the problem with conjoined constraints like the ones used by Lubowicz becomes

apparent when one considers the logical equivalent: a segment should be faithful if it is

marked.[10] Intuitively, such a principle is odd, and indeed, these constraints have the potential to

create typological monsters. Hayes and White illustrate this possibility with a hypothetical

language where voiced obstruents are allowed as the middle consonant in a CCC cluster, but

otherwise are illegal in the language.[11] As Hayes and White put it, we are unlikely to find such a

language, and moreover, it seems undesirable to predict that highly marked sequences would

---

[10] This was previously noted by Ito & Mester (1998), who illustrated the problem with a different hypothetical
example. In the 1998 paper, they argued that conjunction of a markedness constraint with a faithfulness constraint
should not be allowed in the theory.

[11] The relevant constraint ranking for this scenario is [IDENT(voice) & *CCC] >> MAX(C) >> *[-son, +voice] >>
*CCC >> IDENT(voice) with hypothetical inputs like /da/ → [ta] but /apdka/ → [apdka].

license a segment that is otherwise illegal in the language. Without an implemented theory of

how conjoined constraints are learned, it is unclear how the theory could be constrained so that

the problematic cases are avoided while the desired cases are allowed.

2.5.2 *MAP constraints and the P-map

I propose instead to modify the constraint set in a different way, by adopting the family of

*MAP correspondence constraints proposed by Zuraw (2007). Unlike in traditional IDENT

constraints, *MAP constraints are not restricted to penalizing changes in a single feature; instead,

they penalize correspondence between any two natural classes of sounds. The constraints are

formalized as follows, adopted from Zuraw (2007):

(23) ***MAP formalized***
    *MAP(x, y): violated when a sound that is a member of natural class x corresponds to a
    sound that is a member of natural class y.[12]

For the purposes of the cases considered here, what will be necessary is segment-specific

versions of the constraints. For instance, *MAP(p, β) would be violated whenever [p] is in

correspondence with [β].[13] In this case, the constraint *MAP(p, β) may be considered notational

---

[12] Zuraw's formalism also specifies particular contexts in which the pair of sounds must not be in correspondence
(e.g., a sound of natural class x in context A__B should not correspond to sound of natural class y in context C__D).
The context-specific version of the constraints is not necessary here, so I stick to this context-free version for
simplicity.

[13] Hypothetically, any correspondence relationship (i.e., input-output, output-output, base-reduplicant) is possible,
but an input-output correspondence is conceptually odd in this case because the constraints are intended to be
sensitive to the relative similarity of the sounds involved. It is unclear how to judge the similarity between an
abstract input form and a surface form. My analysis of saltation is fully consistent with an output-output
interpretation of the constraints, as is discussed further in section 4.6.3.

shorthand for *Map($\begin{bmatrix} -\text{voice} \\ -\text{cont} \\ +\text{labial} \end{bmatrix}$,$\begin{bmatrix} +\text{voice} \\ +\text{cont} \\ +\text{labial} \end{bmatrix}$), where each of the corresponding natural classes

happens to be made up of only a single segment.

It is worth noting that these *Map constraints are not inconsistent with traditional

faithfulness constraints. Indeed, traditional faithfulness constraints can be treated as special cases

of *Map constraints—for instance, *Map([−voice], [+voice]) would be violated whenever a

voiceless sound is in correspondence with a voiced sound. Likewise, *Map(C, ∅) would be

violated whenever a consonant is in correspondence with zero, making it equivalent to Max-C.

Thus, segment-specific faithfulness constraints and traditional faithfulness constraints can be

unified into the same family of constraints.

Adopting the family of *Map constraints, we see that even (otherwise) classical OT

straightforwardly allows saltation. The solution for the Campidanian Sardinian case, where /p/ →

[β] but /b/ remains [b], is shown in the tableaux in (24). The markedness constraints *V[−cont]V

and *V[−voice]V are ranked above *Map(p, β) so that underlying /p/ will change to [β].

*Map(b, β) can then be ranked above *V[−cont]V so that /b/ is protected from changing.

(24) *Deriving saltation in OT with *Map constraints*

   a. */p/ → [β]*

/VpV/	*Map(b, β)	*V[−cont]V	*V[−voice]V	*Map(p, β)	*Map(p, b)
☞ VβV				*	
VbV		*!			*
VpV		*!	*		

b. */b/ → [b]*

/VbV/	*MAP(b, β)	*V[−cont]V	*V[−voice]V	*MAP(p, β)	*MAP(p, b)
VβV	*!				
☞ VbV		*			

The solution works because the constraint set renders it possible for short "journeys" (i.e., b →
β) to be considered worse than long "journeys" (i.e., p → β).

The reader may be (justifiably) concerned that the addition of *MAP constraints is too
powerful for phonological theory. However, Zuraw's *MAP formalism comes with a substantive
proposal based on Steriade's (2001/2008) P-map theory (for an overview, see section 1.2.2.2),
which serves to constrain the theory. Following Steriade, Zuraw proposes that the knowledge
encoded in the P-map is translated into *a priori* rankings for the *MAP constraints. By default,
*MAP constraints penalizing correspondences between perceptually similar sounds are ranked
lower in the hierarchy than constraints penalizing correspondences between less similar sounds
(e.g., *MAP(p, β) is initially ranked higher than *MAP(b, β)). Of course, in order to have a
saltation this hierarchy must eventually be overturned so that *MAP(b, β) ranks higher than
*MAP(p, β). Under Zuraw's proposal, the P-map only specifies a default ranking, which can be
overturned through learning.

In this dissertation, I adopt Zuraw's theory of *MAP constraints augmented with a
substantive bias based on the P-map as the basis of my analysis of saltation. In Chapter 4, this
theory is formalized within a Maximum Entropy learning model, with the bias implemented
computationally as a prior. We will see that the resulting analysis makes predictions that are
consistent with experimental results (Chapter 3), and more generally, with the overall pattern that
I argue is desirable: that saltations are learnable, but have a dispreferred status.

2.6 Summary of the chapter

In this chapter, I introduced the phenomenon of saltation and provided several examples attested in natural languages. We saw that overall saltation appears to be cross-linguistically rare, and most of the attested cases have quirks that bring into question their status as stable, productive phonological processes in the language. I presented evidence (following Hayes and White, in prep.) that saltation arises by historical accident, and I conjectured (with evidence to follow in coming chapters) that saltation is a dispreferred pattern for language learners. Finally, I demonstrated that classical OT and other closely related frameworks (Harmonic Serialism, Harmonic Grammar) are unable to derive saltation with the traditional set of constraints assumed. After outlining and dismissing Lubowicz's (2002) local constraint conjunction proposal as a solution, I provided a brief overview of an analysis of saltation (which I adopt in Chapter 4) consisting of *MAP faithfulness constraints (from Zuraw, 2007), augmented with a substantive bias based on the P-map (Steriade, 2001/2008).

In Chapter 3, I present evidence from two artificial language experiments with adults indicating that saltatory alternations are indeed dispreferred by language learners.

CHAPTER 3

Experimental evidence that saltatory alternations are dispreferred by adult learners

3.1 Background: Acquisition of phonological alternations

A PHONOLOGICAL ALTERNATION occurs when a morpheme is pronounced differently depending on its phonological context. In American English, for instance, the verb root *pat* is pronounced with a final [t] in the word *pats* [pæts] but with a tap sound [ɾ] in the word *patting* [pæɾɪŋ]. Native speakers of English know that the words *pats* and *patting* are related to the same verb root *pat* even though the root itself is pronounced differently in the two words. More generally, adult speakers tacitly know the distribution of phonological variants in their language and they are able to map multiple surface variants of a lexical item to the same representation at an abstract level (Lahiri & Marslen-Wilson, 1991). Thus, learning the alternations of one's language must be part of the language acquisition process.

Despite their importance in phonological theory, we still know relatively little about when learners acquire the phonological alternations of their language. Several studies have shown that adults are able to learn novel phonological alternations after brief exposure to an artificial language (Onishi, Chambers, & Fisher, 2002; Pycha et al., 2003; Pater & Tessier, 2003, 2005; Peperkamp & Dupoux, 2007; Moreton, 2008; Finley & Badecker, 2009; Skoruppa et al., 2011), as are 12-month-old infants (White et al., 2008).

Our understanding of how this acquisition occurs—that is, which mechanisms are involved in the learning—is even less advanced. One component likely involved in the process of

acquiring phonological alternations is tracking statistical properties of the input. The ability to

track distributions is undeniably a powerful tool available to the language learner. Infants have

been shown to use distributional learning in several aspects of early phonological acquisition,

including discrimination of speech sounds (Maye, Werker, & Gerken, 2002; Anderson, Morgan,

& White, 2003; Yoshida et al., 2010), phonotactic learning (Chambers, Onishi, & Fisher, 2003),

and word segmentation (Saffran, Aslin, & Newport, 1996). Adults have exhibited a similar

ability to use statistical learning when segmenting novel words in an artificial language (e.g.,

Saffran, Newport, & Aslin, 1996).

A plausible starting point for learning alternations is by looking for complementary

distributions among speech sounds, that is, by looking for cases where two speech sounds never

occur in the same phonological environment (e.g., Peperkamp, Le Calvez, Nadal, & Dupoux,

2006a).[14] For instance, infants exposed to English may notice that [t] and [ɾ] never occur in the

same environment, leading them to analyze the two sounds as alternating variants of the same

phoneme.

However, this process is unlikely to be based on distributional information alone. In

English, for instance, the sounds [h] and [ŋ] happen to have completely non-overlapping

distributions because [h] only occurs at the beginning of syllables (as in [hæt] *hat*) and [ŋ] only

occurs at the ends of syllables (as in [sɪŋ] *sing*). No phonological analysis, however, would claim

that [h] and [ŋ] are context-dependent variants of the same underlying sound because, other than

---

[14] Peperkamp et al.'s model was actually designed to learn allophonic rules, not alternations per se. Allophones are variants of a phoneme that occur only in particular phonological contexts (e.g., [pʰ] and [p] are allophones of the phoneme /p/: [pʰ] occurs at the beginning of syllables, unless preceded by an [s], in which case [p] appears). Allophones result in alternations when the same morpheme variably appears with multiple allophones depending on context (e.g., *pat* may end in [t] and [ɾ]), but some pairs of allophones have only distributional evidence. In this discussion, I am assuming that there are some mechanisms used for learning all allophones, regardless of whether they result in alternations. Those that result in alternations, of course, have paradigmatic evidence of their relationship that purely distributional allophones do not have. However, any differences in how allophonic relationships with and without supporting evidence from alternations are learned remain poorly understood.

being consonants, the two sounds are phonetically distinct from each another in almost every possible way (see Peperkamp et al., 2006a for a similar case in French).

Here, I argue that learners consider the similarity of the sounds involved when learning phonological alternations, which helps them from settling on erroneous phonological mappings. This idea is consistent with Steriade's theory of the P-map (discussed in section 1.2.2.2). Steriade proposed that learners are biased to expect that phonological processes will involve minimal modification. In this study, I test this proposal by looking at saltation, a pattern that represents not minimal modification, but excessive modification.

3.2 Saltatory alternations: A case of excessive modification

A SALTATORY ALTERNATION refers to a phonological alternation in which an intermediate, non-alternating sound must be "leaped over." An illustrative example of a saltatory alternation comes from Campidanian Sardinian (Bolognesi, 1998). In this language, voiceless stops [p, t, k] become voiced fricatives [β, ð, ɣ] after vowels (as in 25a), but voiced stops [b, d, g] remain unchanged in that context (as in 25b). Crucial sounds are denoted with bold font:

(25)   *Example of a saltatory alternation in Campidanian Sardinian (Bolognesi, 1998)*

Isolation Form		Post-vowel Form	Meaning
a. [**p**ãi]	→	[sːu **β**ãi]	'the bread'
[**t**rintaduzu]	→	[sːu **ð**rintaduzu]	'the thirty-two'
[**k**uatːru]	→	[dɛ **ɣ**uatːru]	'of four'
b. [**b**ĩu]	→	[sːu **b**ĩu],	'the wine'
[**d**ɔmu]	→	[dɛ **d**ɔmu]	'of house'
[**g**ɔma]	→	[dɛ **g**ɔma]	'of rubber'

For a more in depth introduction to saltation, including a careful definition and several examples from real languages, see Chapter 2.

Saltatory alternations, like the ones in Campidanian Sardinian, represent striking counterexamples to the principle of minimal modification. The fact that [t], for instance, is changed when it occurs after vowels is presumably driven by a ban on post-vowel voiceless stops in Campidanian Sardinian. The change from [t] to [ð] already represents an alternation between sounds that are not minimally different, but this alternation would not be particularly troubling if the language also banned post-vowel voiced stops (e.g., [d]). However, because intermediate [d] appears unchanged after vowels in Campidanian Sardinian (i.e., it does not alternate), the alternation between [t] and [ð] represents excessive modification. It is unclear why [t] changes into [ð] when changing instead to [d], which is legal in that context, would require a less extreme modification. Looking at it from a different angle, if speakers of Campidanian Sardinian tolerate an alternation between sounds as dissimilar as [t] and [ð], why would they not tolerate an alternation between similar sounds, such as [d] and [ð]? Intuitively, saltatory alternations represent a contradiction to the principle of minimal modification because long journeys are allowed but short journeys are not. Indeed, due to this atypical characteristic, classical Optimality Theory (OT; Prince & Smolensky, 1993/2004) predicts that saltatory alternations should not exist (section 2.4; see also Lubowicz, 2002; Ito & Mester, 2003).

If learners are biased to assume that phonological changes will be minimal, as claimed by Steriade (2001/2008), then the excessive nature of saltatory alternations should cause them to be dispreferred by learners. I test this hypothesis using the artificial language learning paradigm.

51

3.3 Overview of the experiments

The experiments each consisted of three phases: exposure, verification, and generalization. In the exposure phase, participants learned alternations by listening to pairs of nonwords representing singular and plural nouns in an artificial language, paired with pictures of singular and plural items, respectively. Plural words were always formed by adding [i] to the end of singular words. The singular words ended in target sounds, some of which changed when the [i] suffix was added in the plural word (e.g., singular [kamap], plural [kamavi]), providing the basis for the phonological alternations that participants were learning. Participants were also exposed to examples in which the target sound was a filler sound that did not change in the plural form (e.g., singular [luman], plural [lumani]).

In the verification phase, participants were tested on <u>a subset of words from the exposure phase</u> using a two-alternative forced-choice task. Participants heard a singular word and then chose between two possible plural forms, one with a changed final target sound and one with an unchanged target sound. For each trial, one of the plural options (either the changing or non-changing option) followed the pattern learned during exposure and the other option was a foil that did not follow the learned pattern.

The generalization phase was similar to the verification phase, except participants were tested on <u>novel words</u> that were not presented during the exposure phase. Some of the novel words ended in the same target sounds from the exposure phase, but to test for an inductive bias, a subset of the novel words in the generalization phase contained new target sounds that were not presented during exposure.

3.4 Experiment 1: Potentially saltatory input

Experiment 1 was designed to test an implicational question: given that a learner has acquired a *potentially* saltatory alternation, which assumptions does the learner make about untrained, intermediate sounds? Participants were randomly assigned to two conditions, Potentially Saltatory or Control. Figure 3 summarizes the input provided during the exposure phase for each of these conditions. Participants in the Potentially Saltatory condition were trained on alternations between voiceless stops and voiced fricatives ([p ~ v] and [t ~ ð]) during the exposure phase, but crucial examples of the intermediate consonants [b, f, d, θ] were withheld. The alternations learned during exposure in the Potentially Saltatory condition were thus ambiguous: they would be saltatory if intermediate sounds remained unchanged, but non-saltatory if intermediate sounds also alternated with voiced fricatives. In the Control condition, participants were instead trained on the alternations [b ~ v] and [d ~ ð] with examples of [p, f, t, θ] withheld. The alternations in the Control condition were unambiguously non-saltatory because none of the withheld sounds were intermediate between the alternating sounds.

Figure 3. Summary of the input during the exposure phase and possible interpretations of the input for the Potentially Saltatory and Control conditions in Experiment 1.

Potentially Saltatory condition:

Training input:

(a)   Labials   Coronals

    p   t

         v   ð

Possible interpretations of the input:

(b) Saltatory, No new alternations posited

(c) Non-saltatory, New alternations posited

Control condition:

Training input:

(d)   Labials   Coronals

        b   d

        v   ð

Possible interpretations of the input:

(e) Non-saltatory, No new alternations posited

(f) Non-saltatory, New alternations posited

The Control condition acted as the baseline for comparison in this experiment. As Figure 3 demonstrates, participants in the Control condition could choose to treat the untrained sounds as alternating (3f) or non-alternating (3e), but in both cases, the system learned would be non-

54

saltatory. Assuming that learners are generally reluctant to extend patterns to unseen sounds (e.g., Peperkamp & Dupoux, 2007), participants were predicted to change untrained sounds relatively infrequently in the Control condition because they had no evidence for doing so. In the Potentially Saltatory condition, participants had the same choice between changing untrained sounds without evidence (3c) or leaving them unchanged (3b). However, changing untrained sounds in the Potentially Saltatory condition (unlike in the Control condition) would avoid a saltatory alternation. Thus, if learners disfavor saltatory alternations, participants should change untrained sounds more often in the Potentially Saltatory condition than in the Control condition. In other words, saltation avoidance should lead participants in the Potentially Saltatory condition to counteract any default inclination (relevant to both conditions) towards being conservative.

3.4.1 Method

3.4.1.1 Participants

Forty undergraduate students in introductory psychology or linguistics classes at UCLA completed the experiment for partial course credit. Seven additional participants (2 in the Potentially Saltatory condition, 5 in the Control condition) began the experiment but did not complete it because they failed to reach the criterion in the verification phase within the allotted time (see section 3.4.1.3 below). These participants received credit, but their data were not used in the analysis.

3.4.1.2 Materials and apparatus

Exposure phase. For the exposure phase, 72 nonwords of the form CVCVC (e.g., [kamap]) were created as singular stimuli for the Potentially Saltatory condition. Half of the nonwords ended in

55

the target sounds {p, t}, 18 of each, and half of the nonwords ended in one of the filler sounds {m, n, l, ɹ, s, ʃ}, 6 of each. The initial consonant sounds were drawn from the set {p, b, t, d, k, g, f, θ, s, ʃ, m, n, l, ɹ}. Because the crucial context for the alternations was between vowels, the medial consonants were chosen from the more limited set of filler sounds {m, n, l, ɹ, s, ʃ} so that the middle consonants would not provide unintended distributional information. Vowels were drawn from the set {i, a, u}. Nonwords were created by combining the possible consonants and vowels for each slot in a pseudorandom manner. Each consonant and vowel was used an approximately equal number of times in any given word position with the exception of the word-final position, which followed the proportions described above. Resulting nonwords were thrown out and replaced if they closely resembled real English words as judged by a native speaker (the author) or if they contained the same consonant in all positions (e.g., [ʃuʃuʃ]).

For each of the 72 singular nonwords, a plural form was also created. For nonwords ending in fillers sounds, plural forms were created by adding the vowel [i] to the end of each singular nonword with no change in the final consonant (e.g., singular [luman], plural [lumani]). For nonwords ending in {p, t}, a final [i] was added and the final consonant was changed to the corresponding voiced fricative, either [v] or [ð] (e.g., singular [kamap], plural [kamavi]).[15]

Stress was placed on the second syllable of all words, that is, on the final syllable of CVCVC singular words and on the middle syllable of CVCVC-i plural words. This pattern, consistent with a stress system in which stem-final vowels receive stress (e.g., as attested in

---

[15] For the coronal sounds, [ð] was chosen as the voiced fricative rather than logically possible [z] for two reasons. First, [z] changes the extra phonological feature [strident], which is not changed by [ð]. Second, [ð] was chosen to remain as close as possible to the attested case in Campidanian Sardinian (described in section 2.1) while still using sounds that are English phonemes, which should be more easily distinguished by the English-speaking participants. The main predictions related to saltation avoidance still hold if [s] and [z] had been used instead of [θ] and [ð], although the precise numerical outcomes may have been slightly different due to differences in the perceptual similarity of the sounds involved (e.g., see the modeling work in Chapter 4).

Albanian, see Bevington, 1974, p. 24), was chosen so that stress would be on the same syllable of the stem in the singular and plural forms.

For the singular nonwords ending in {p, t}, corresponding nonwords for the Control condition were created by changing each final [p] to [b] and each final [t] to [d]. The same list of 36 singular nonwords ending in filler sounds from the Potentially Saltatory condition was used in the Control condition without alteration. Plural forms for the Control condition were created in the same manner described above. Thus, the list of stimuli for the Potentially Saltatory condition and the Control condition differed only in the final target sound of the non-filler items. For example, singular [kamap] and plural forms [kamapi] and [kamavi] in the Potentially Saltatory condition corresponded to singular [kamab] and plural forms [kamabi] and [kamavi] in the Control condition.

Each set of nonwords was randomly paired with one of 72 pictures showing singular objects (e.g., a strawberry) and 72 corresponding pictures showing multiple objects (e.g., two strawberries). The pictures were made up of clipart-style images or small photographs of everyday nouns taken from the Internet. The number of objects in the plural pictures was always greater than one, but otherwise varied. Corresponding nonwords in the Potentially Saltatory condition and the Control condition were paired with the same pictures.

Verification phase. For the Potentially Saltatory condition, 32 of the singular nonwords (8 *p*-final, 8 *t*-final, and 16 fillers), along with their associated pictures, were randomly chosen from the set of nonwords in the exposure phase for use in the verification phase. For each nonword in this phase, it was necessary to have both a changing plural option and a non-changing plural option. To make a changing plural option for the singular nonwords ending in filler sounds, the

following correspondences were used: [m, ʃ, ɹ] changed to [v], and [n, s, l] changed to [ð]. As examples, singular [kamap] had plural options [kamapi] and [kamavi], and singular [luman] had plural options [lumani] and [lumaði]. In the Control condition, the corresponding set of nonwords was used, and the plural options were formed in the same manner.

Generalization phase. For the generalization phase, 72 new singular nonwords were created in the same manner described above. For the Potentially Saltatory condition, one-third ended in {p, t} (12 of each), one-third ended in the filler sounds {m, n, l, ɹ, s, ʃ} (4 of each), and one-third ended in the intermediate sounds {b, d, f, θ} (6 of each). For the Control condition, the same set of words were used except word-final [p] was changed to [b], word-final [t] was changed to [d], and vice versa. Thus, one-third of the Control nonwords ended in {b, d}, one-third ended in the filler sounds {m, n, l, ɹ, s, ʃ}, and one-third ended in the sounds {p, t, f, θ}. Changing and non-changing plural forms were created in the same manner described above. The nonwords for the generalization phase were randomly assigned to 72 new pairs of pictures showing singular and plural objects. A sample of the nonwords used is provided in the Appendix (section 3.8).

Stimuli recording and experimental apparatus. A male native speaker of English with phonetic training, who was unaware of the purpose of the experiment, recorded the nonwords in a soundproof booth using a Shure SM10A head-mounted microphone, whose signal ran through an XAudioBox pre-amplifier and A-D device. The recordings were done using PcQuirerX at a sampling rate of 22,050 Hz. In order to make the relevant contrasts as perceptible as possible, a relatively careful speech style was used. The speaker was asked to release all word-final consonants and to fully voice all voiced segments. The spectrogram for each token was

58

inspected using Praat (Boersma & Weenink, 2010) to confirm that voicing and frication was present in voiced sounds and fricatives, respectively.

Stimuli for the Potentially Saltatory condition and the Control condition were recorded on separate occasions, but every effort was made to ensure that the stimuli were produced in a consistent way. To ensure that the target sounds were perceived equally well in the two conditions, five native speakers of English listened to the singular nonwords ending in the target sounds [p, t, b, d, f, θ] from the generalization phase and the corresponding changing and non-changing plural options (288 total nonwords). They were asked to identify the target sound (i.e., the last consonant) for each nonword by choosing one of eight options, [p, t, b, v, d, f, θ, ð]. The sounds [θ] and [ð] were described as "<th> in *thick*" and "<th> in *the*", respectively. Accuracy was very high overall, and crucially, it was comparable between the Potentially Saltatory condition (94.3%) and the Control condition (94.7%), indicating that the target sounds were perceived equally well across the two conditions.

The experiment was conducted in a quiet room on a Dell computer equipped with a 20-inch monitor and Sony MDR-V200 headphones. The experimental software E-prime (version 2.0) was used to present the stimuli and record the responses.

3.4.1.3 Procedure

The experiment consisted of three phases: exposure, verification, and generalization. In the exposure phase, participants were instructed that they would be learning words in a foreign language. They were told that they should try their best to remember the words because they would be tested on them later. Participants were told to repeat each word out loud after hearing it because doing so would help them remember. Participants heard 72 unique, self-paced trials in

this phase. Each trial began with a picture showing a singular object appearing in the center-left part of the computer screen. After the picture had been displayed for one second, the singular nonword for that item was played over headphones. The singular picture disappeared 2.5 seconds after the sound file began playing, and the corresponding plural picture immediately appeared in the center-right part of the screen. The plural nonword for that picture was played over headphones one second after the plural picture appeared. After hearing both nonwords, participants pressed the spacebar, which initiated the next trial. The plural picture remained on the screen until the participant pressed the spacebar. Nonwords were only presented in auditory form, never in orthography. Participants were given no further instructions. The order of trials in this phase, as well as in the following two phases, was randomized anew for each participant by E-prime. The exposure phase lasted approximately 10-25 minutes depending on how quickly a given participant pressed the spacebar.

In the verification phase that followed, participants were tested on 32 words that they had heard during the exposure phase. The purpose of the verification phase was to ensure that participants had successfully learned the pattern presented during exposure. A singular picture appeared on the left side of the screen and the singular nonword for that picture was played over headphones. Once the singular picture disappeared, the plural picture was displayed on the right side of the screen along with a row of question marks located just under the picture. Up to this point, the trial was identical to trials in the exposure phase (except for the question marks), including the timing of the stimuli. After the plural picture was on the screen for 1.5 seconds, participants heard two plural options—the changing plural option and the non-changing plural option—with a one second pause in between them. Order of the two plural options was counterbalanced such that the changing option and the non-changing option occurred first an

equal number of times for each type of singular word. Participants were asked to choose the correct word for the plural picture by pressing the appropriate key on the keyboard: a key marked "1" for the first option or a key marked "2" for the second option (the "f" and "j" keys, respectively, were used for this purpose). The next trial started immediately after a response key was pressed. The verification phase lasted approximately five minutes. At the end of the phase, a screen appeared showing the participant's accuracy. If participants did not achieve an accuracy of at least 80%, they were told that they needed to reach 80% to continue in the experiment. They then repeated the exposure and verification phases until they reached an accuracy of at least 80%. Participants heard the same trials on subsequent exposure and verification phases, but in a different random order. Participants continued cycling through the exposure phase and verification phase until they either reached 80% or 50 minutes had elapsed (typically after two cycles). Those who did not reach 80% after 50 minutes did not complete the generalization phase.

Participants who achieved 80% accuracy on the verification phase moved into the generalization phase, where they were tested on 72 novel words, including words ending in untrained target sounds. Otherwise, trials were identical to those in the verification phase. Participants were instructed that they would be hearing new words in the same language and that they should make their best guess based on their experience so far with the language. The generalization phase lasted approximately 10 minutes.

3.4.2 Results

Only responses in the generalization phase (i.e., responses to words not encountered during exposure) were included in the analysis. The data were analyzed using mixed-effects logistic

regression models (see Jaeger, 2008), implemented in R (R Core Development Team, 2008) using the *lme4* package (Bates, Maechler, & Dai, 2008). To compare models, likelihood ratio tests were conducted using the anova() function. The likelihood ratio test compares the log likelihoods of two models with different numbers of factors (in a subset relationship) and determines if the added factors are justified based on a chi-squared test (see Baayen, 2008, ch. 7). The random effect structures of the models were determined by backwards stepwise comparison, that is, by taking out each random effect factor one at a time, comparing the simpler model to the more complex model using likelihood ratio tests, and removing the random factor if it did not significantly improve model fit (see Baayen, Davidson, & Bates, 2008).

3.4.2.1 Trained sounds

I first consider how well participants extended the patterns learned during the exposure phase (i.e., change [p, t] to [v, ð] in the Potentially Saltatory condition or [b, d] to [v, ð] in the Control condition, and do not change fillers sounds) to new words of the same type in the generalization phase. There was no reason to expect a difference between the Potentially Saltatory condition and the Control condition on trained sounds because participants in both conditions had to reach the 80% criterion in the verification phase to move on to the generalization phase. Indeed, the results show that accuracy on trained sounds was similar in the Potentially Saltatory condition (93.0%) and the Control condition (92.2%). A mixed logit model predicting log odds of an accurate response, with random intercepts for subjects and a fixed effect of Condition (Potentially Saltatory condition vs. Control condition), found that Condition was not a significant predictor ($p = .76$), and including it in the model did not significantly improve model fit, $\chi^2(1) = .09$, $p = .76$. As evidenced by the mean accuracies above 90%,

participants in both conditions readily extended the trained patterns to new words ending in the same sounds.

3.4.2.2 *Untrained sounds*

The primary purpose of Experiment 1 was to see how participants would treat untrained, intermediate sounds given that they had learned a potentially saltatory alternation. Recall that if learners have a bias that disfavors saltatory alternations, participants were predicted to change untrained sounds more often in the Potentially Saltatory condition than in the Control condition. Figure 4 shows the results for untrained sounds according to Condition and Sound Type (untrained stops vs. untrained fricatives). Because participants received no information about words ending in these sounds during the exposure phase, there was no "correct" answer, so accuracy cannot be calculated. Instead, the mean percent of trials in which participants chose the changing plural option is reported (e.g., for the singular word [talab], how often did they choose [talavi] rather than [talabi]). Overall, we see that participants tended to change the untrained intermediate sounds more often in the Potentially Saltatory condition than in the Control condition for both untrained stops (70.0% vs. 20.8%) and the untrained fricatives (45.0% vs. 15.8%). Within the Potentially Saltatory condition, participants also showed a tendency to change untrained stops more often than untrained fricatives (70% vs. 45%).

To evaluate these differences, a mixed logit model was fitted, predicting log odds of having a changing response for words ending in untrained target sounds. The final model included fixed effects for Condition (Potentially Saltatory vs. Control), Sound Type (stops vs. fricatives), and a Condition x Sound Type interaction. Random intercepts for subjects and by-subject random slopes for Sound Type were also included. By-subject random slopes were included because

they significantly improved model fit according to a likelihood ratio test, $\chi^2(3) = 75.62$, $p < .001$. Random intercepts for individual words were not included in the final model because they did not significantly improve model fit, $\chi^2(1) = .12$, $p = .72$.

The fixed effects for the final model are provided in Table 2. The significant negative intercept indicates that words in the Control condition (which acts as the baseline in this model) were changed infrequently overall, suggesting that learners are conservative when they encounter untrained items. Condition was a significant predictor in the model, indicating that participants chose the changing option for words in the Potentially Saltatory condition (i.e., those with final intermediate sounds) significantly more often than for words in the Control condition. These results are consistent with the main prediction: participants changed untrained sounds more often when they were intermediate between alternating sounds. There was also a significant interaction, indicating that untrained stops were changed more frequently than unchanged fricatives, but only in the Potentially Saltatory condition.

Figure 4. Results for untrained sounds in Experiment 1 by Condition and Sound Type. Individual results (diamonds) and overall means (bars) are provided.
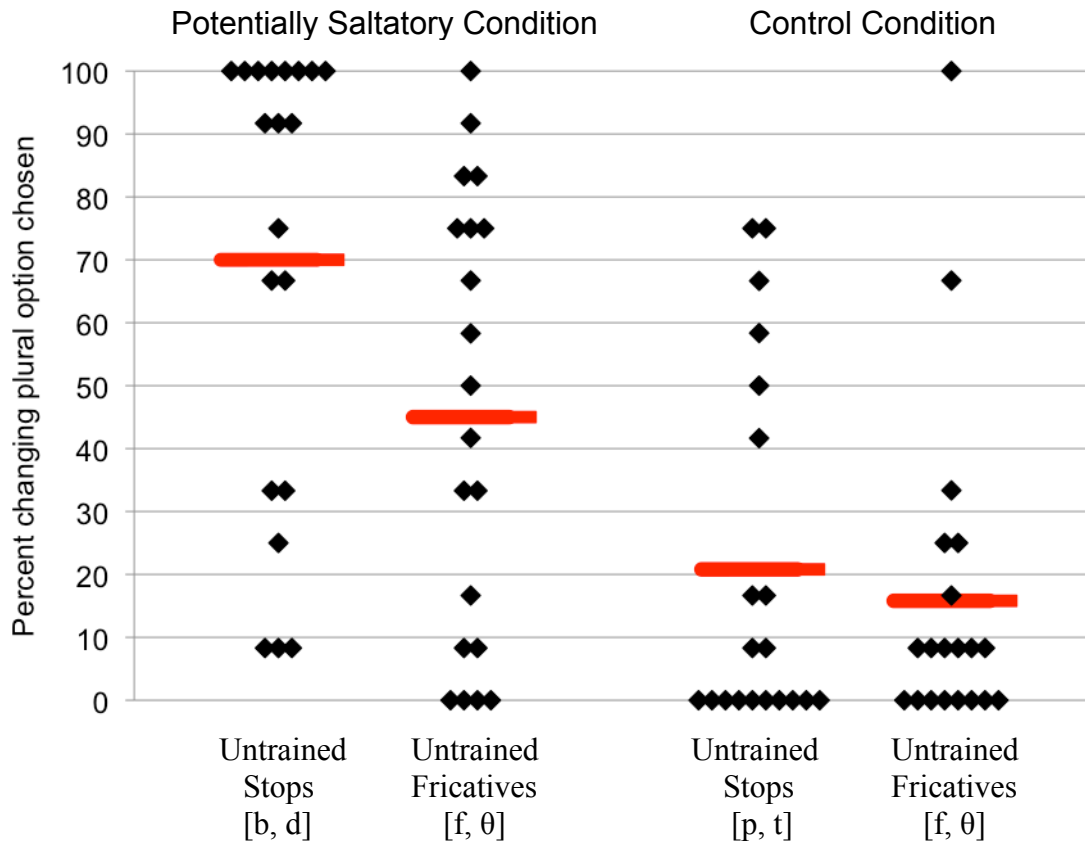
Table 2. Summary of the fixed effects for untrained sounds in Experiment 1.

Predictor	Estimate	Standard error	z-value	p-value
Intercept	−2.80	.57	−4.87	<.001
Condition = *Potentially Saltatory*	2.35	.78	3.02	.002
Sound Type = *Untrained stops*	− .33	.72	− .46	.65
Interaction = *Potentially Saltatory & Untrained stops*	2.80	.97	2.89	.004

3.4.2.3 Effect of amount of exposure

Due to the experimental design, participants received variable amounts of training, either completing one or two cycles of the exposure phase.[16] This design choice was made because of the implicational nature of the hypothesis: given that a participant <u>has learned</u> a potentially saltatory alteration, how does the participant treat untrained, intermediate sounds? To answer this question, it was more critical to ensure that participants had actually learned the potentially saltatory alternations (or the comparable non-saltatory alternations in the Control condition) before being tested on new cases, as opposed to ensuring that all participants received the same amount of exposure. However, it is possible that the amount of exposure, rather than the intermediate status of the untrained sounds, can explain the differences observed between the Potentially Saltatory condition and the Control condition.

To address this possibility, a new mixed logit model for untrained sounds was run with an added fixed effect for number of cycles through the exposure phase (either one or two). In the model, amount of exposure was not a significant predictor ($p = .46$) and its inclusion did not significantly improve model fit according to a likelihood ratio test, $\chi^2(1) = .50$, $p = .48$. An additional model was run with interaction effects between amount of exposure and each of the other fixed effects. None of the effects related to amount of exposure reached significance in the model. Moreover, including the added fixed effects for amount of exposure and the associated interactions did not significantly improve model fit, $\chi^2(4) = 5.19$, $p = .27$. In both models, the effect of Condition (Potentially Saltatory vs. Control) remained significant ($p = .01$). Overall,

---

[16] In principle, it was possible to have three cycles, but no participant who completed Experiment 1 within the allotted hour had more than two cycles of the exposure phase.

these models indicate that the amount of exposure did not have a significant effect on how often participants chose the changing option for untrained sounds in Experiment 1.

3.4.3 Discussion

The results from Experiment 1 were consistent with the idea that learners have a general preference for saltation avoidance. Participants who learned potentially saltatory alternations during the exposure phase (Potentially Saltatory condition) changed intermediate sounds at a high rate despite having no direct evidence for such changes in the input. By changing intermediate sounds, participants avoided the dispreferred saltatory alternations that would result if the intermediate sounds remained unchanged. Participants in the Control condition learned comparable alternations, but were not under the same pressure to avoid saltation. As predicted, they changed the untrained sounds much less frequently than participants in the Potentially Saltatory condition. The results can also be summarized in the following terms: participants learning alternations between dissimilar sounds extended the pattern to alternations between more similar sounds, consistent with the principle of minimal modification.

The Control condition served as an important baseline for comparison because it provided us with an idea of how often participants would change untrained sounds when saltation was not a factor. We see that participants were not making responses at random for the non-intermediate untrained sounds in the Control condition, which would have resulted in chance performance (50%). Rather, participants appear to have sensibly taken the more conservative approach, that is, they were reluctant to posit new alternations without evidence, consistent with previous findings (e.g., Peperkamp & Dupoux, 2007). The relatively high rate that participants chose the changing option in the Potentially Saltatory condition is even more striking compared to the low

rate that the changing option was chosen in the Control condition. Together, the results show that the saltation avoidance effect was strong enough that learners were willing to go against their default preference to avoid positing new alternations without evidence.

The Control condition also rules out two possible alternative explanations for why participants in the Potentially Saltatory condition might have changed untrained sounds. First, participants may have been giving product-oriented responses (Bybee & Slobin, 1982). In other words, participants may have responded based on the form of the *product* (i.e., the plural word) rather than based on the form of the singular word. During exposure, participants heard a large proportion of plural words with either [v] or [ð] as the final consonant. In fact, half of the plural forms had [v] or [ð] as the final consonant whereas only one-twelfth of the total plural forms in the exposure phase had any one of the other possible final consonants (see section 3.4.1.2). Participants may have responded to any untrained cases by matching the frequency of the plural endings that they heard during training. This strategy would have resulted in a preference for the changing option for any novel sound. However, the same proportion of changing and non-changing plural forms was heard during the exposure phase in both the Potentially Saltatory condition and the Control condition. Thus, if participants were using (only) a product-oriented strategy, they should have changed an equal percentage of untrained sounds in both conditions, contrary to the results.

A second possibility is that participants may have been biased to target a more general class of phonetically similar sounds (e.g., Saffran & Thiessen, 2003; Skoruppa & Peperkamp, 2011). In particular, they may have learned that all stops (or all obstruents) changed to voiced fricatives between vowels, rather than limiting the alternations only to voiceless stops. It has been argued that phonological generalizations are easier to learn when they can be described using fewer

features (e.g., Skoruppa & Peperkamp, 2011; Moreton & Pater, 2012a). Because examples ending in voiced stops and voiceless fricatives were withheld from training, targeting a more general class of sounds would be equally consistent with the input, would require fewer features for grouping the sounds, and would explain the extension to untrained sounds in the Potentially Saltatory condition. However, if participants were only biased to target general classes of similar sounds, it is left unexplained why they did not generalize to untrained sounds (at least to voiceless stops) at comparable rates in the Control condition. Because participants in the Control condition did not change untrained sounds at the same rate as those in the Potentially Saltatory condition, we can conclude that targeting a more general class of sounds cannot fully explain the results in Experiment 1. I return to a more nuanced version of the featural complexity account in section 3.6.2.

Crucially, neither of these alternative explanations—product-oriented responding or targeting general classes of sounds—can account for the difference observed between the Potentially Saltatory condition and the Control condition.   However, it is worth noting that participants did sometimes choose the changing option for untrained sounds in the Control condition (roughly 15–20% of the time). If participants are truly averse to positing new alternations without a reason, then we might expect this value to be very close to 0%. The low, but non-zero, values in the Control condition may simply be due to random noise, but it could indicate as well that the alternate explanations described above had some degree of influence on the responses, at least for some participants (see individual data points in Figure 4). The changing responses in the Control condition may also be due, in part, to task demands. In the two-alternative forced-choice task, some participants may have been motivated to use both response options for untrained sounds, even in the Control condition.

A final noteworthy aspect of Experiment 1 is that participants in the Potentially Saltatory condition preferred to change untrained stops more frequently than untrained fricatives. In terms of phonological features, these sounds are equally different from each other: voiced stops and voiced fricatives differ in one feature ([continuant]), and voiceless fricatives and voiced fricatives differ in one feature ([voice]).[17] If participants prefer alternations between similar sounds compared to alternations between less similar sounds (i.e., if they follow the principle of minimal modification), these results imply that abstract features may not provide a sufficient measure of similarity.[18] I return to this issue in the General Discussion (section 3.6).

Overall, Experiment 1 demonstrated that learners disprefer saltatory alternations when presented with ambiguous input. They changed intermediate sounds, without evidence, so that the alternations learned during exposure were rendered non-saltatory. Experiment 2 was designed to look further at the strength of the anti-saltation bias observed in Experiment 1. Does the bias appear only when the input is ambiguous, or does it endure even when there is unambiguous evidence during training for saltatory alternations? In Experiment 2, participants were exposed to the same alternations as in Experiment 1, but they received explicit evidence that intermediate sounds did not change.

---

[17] That is, the changes are equal if the features for place of articulation are construed broadly as labial or coronal. If labio-dentality is considered part of the feature system to differentiate between the slightly different places of articulation of [b] and [v], then /b/ → [v] requires changing two features, whereas /f/ → [v] requires only one (voicing). Under the similarity bias account (or the complexity account), this would predict a preference to change /f/ → [v] more often, the opposite of the actual results. The same argument holds for the subtle place of articulation distinction for the coronal stops [t, d], which are alveolar, and the coronal fricatives [θ, ð], which are dental (see Hayes, 2009).

[18] An alternative explanation for the difference in untrained stops and fricatives in the Potentially Saltatory condition is that two of the non-changing filler sounds were [s] and [ʃ], which are voiceless fricatives like untrained [f] and [θ]. Even though [s] and [ʃ] are sibilants (unlike [f] and [θ]), it is possible that some participants were reluctant to change [f] and [θ] because they form a natural class of voiceless fricatives with two of the filler sounds. To address this possibility, a version of the Potentially Saltatory condition was run with only sonorants [m], [n], [l], [ɹ] as filler sounds (otherwise identical). A difference was once again found between untrained stops (65% changed) and untrained fricatives (47% changed), indicating that the difference was not due to the presence of [s] and [ʃ] in the set of filler sounds. Nevertheless, [s] and [ʃ] were not included in the set of filler sounds in Experiment 2.

3.5 Experiment 2: Explicitly saltatory input

In Experiment 2, participants were randomly assigned to two conditions: a Saltatory condition and a Control condition. Participants in the Saltatory condition learned the same alternations as in the Potentially Saltatory condition in Experiment 1 (i.e., [p ~ v] and [t ~ ð]). This time, however, they also had cases of non-changing intermediate voiced stops [b, d] (Stops sub-group) or non-changing intermediate voiceless fricatives [f, θ] (Fricatives sub-group) during the exposure phase. As a result, the alternations were explicitly saltatory because there was evidence for an intermediate non-alternating sound. The Stops sub-group received no information about intermediate fricatives [f, θ] during exposure, and the Fricatives sub-group received no information about intermediate stops [b, d]. Participants in the Control condition learned the same alternations as in the Control condition of Experiment 1 (i.e., [b ~ v] and [d ~ ð]), but they too were trained that additional sounds did not alternate. The Stops sub-group had examples of non-changing [p, t], and the Fricatives sub-group had examples of non-changing [f, θ]. The input for each of the four resulting sub-groups (Saltatory/Stops, Saltatory/Fricatives, Control/Stops, and Control/Fricatives) is summarized in Figure 5.

The goal of Experiment 2 was to see if participants would find it difficult to learn (and remember) that intermediate sounds did not alternate, even with explicit evidence in exposure. Provided that the anti-saltation bias observed in Experiment 1 is sufficiently strong, participants in the Saltatory condition were predicted to make more errors on intermediate sounds (i.e., by incorrectly choosing the changing plural option) relative to the number of errors made by participants in the Control condition on comparable sounds that were not intermediate.

71

Figure 5. Summary of input in Experiment 2.

Saltatory condition:
Training input:

Stops sub-group   Fricatives sub-group

p   b   t   d | p   t

v   ð | f   v   θ   ð

Control condition:
Training input:

Stops sub-group   Fricatives sub-group

p   b   t   d | b   d

v   ð | f   v   θ   ð
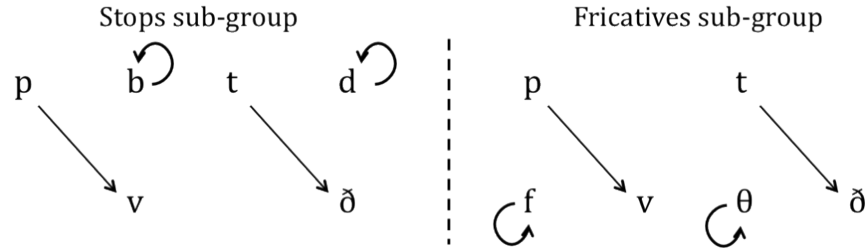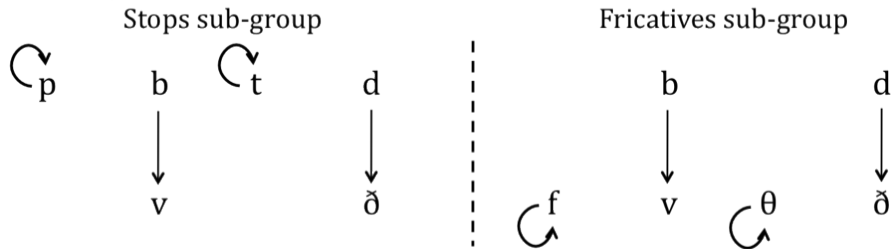
3.5.1 Method

3.5.1.1 Participants

Eighty undergraduate students in introductory psychology or linguistics classes at UCLA completed the experiment for partial course credit. None of the participants had participated in Experiment 1. Twenty-one additional participants (13 in the Saltation condition, 8 in the Control condition) began the experiment but did not complete it because they failed to reach the 80% criterion in the verification phase within the allotted time. These participants received credit, but their data were not used in the analysis. In addition, the data of four participants (all from the Saltation condition) who completed the experiment were not used in the analysis because the participants had clearly not learned (or retained) at least one of the trained alternations (all had

accuracy of less than 10% for one of the trained alternations in the generalization phase).[19] These

participants were replaced by four new participants.

3.5.1.2 Materials

Exposure phase. The exposure phase consisted of 72 singular nonwords. In the Saltatory

condition, half of the singular nonwords ended in the target sounds {p, t}, 18 of each type. Half

of the remaining nonwords (18) ended in the intermediate stops [b, d] (Stops sub-group) or the

intermediate fricatives [f, θ] (Fricatives sub-group), 9 of each type. The final quarter of the

nonwords ended in filler sounds, consisting of {m, n, l, ɹ}, 3–4 of each type. The singular

nonwords were generated as described in Experiment 1. Changing and non-changing plural

forms were created for each of the singular nonwords in the same way described for Experiment

1. The Control condition was analogous, except all target [p, t] sounds were substituted for [b, d],

and vice versa. The same pairs of pictures from Experiment 1 were used in Experiment 2.

Verification phase. For the verification phase, 32 of the nonwords (8 *p*-final, 8 *t*-final, 4 *b*-final

(Stops sub-group) or *f*-final (Fricatives sub-group), 4 *d*-final (Stops sub-group) or *θ*-final

(Fricatives sub-group), and 2 ending in each of the four filler sounds) were chosen at random

from the set of forms used in the exposure phase for the Saltatory condition. The corresponding

set of words was used in the Control condition.

---

[19] The focus of the experiment was to determine how learners would perform on sounds that were intermediate in a saltatory alternation. These four participants clearly had not internalized the alternations that would make the target sounds intermediate, so their data were not relevant to the question in this study.

Generalization phase. For the generalization phase, 64 new singular forms were created (24 ending in [p, t], 12 ending in [b, d] (Stops sub-group) or [f, θ] (Fricatives sub-group), 12 ending in filler sounds, and 16 ending in untrained sounds [f, θ] (Stops sub-group) or [b, d] (Fricatives sub-group)). The Control condition was analogous, except all target [p, t] sounds were substituted for [b, d], and vice versa.

Except as noted, the nonwords for each phase were created and recorded following the same procedure described for Experiment 1. Five native English speakers identified the target sounds of singular nonwords ending in {p, b, f, t, d, θ} and the corresponding changing and non-changing plural options (the task and the five speakers were the same as in Experiment 1, see section 3.4.1.2). Accuracy was comparably high in the Saltatory condition (97.0% correct) and the Control condition (96.5% correct), indicating that the target sounds were perceived equally well in the two conditions.

3.5.1.3 Procedure

The procedure was identical to the procedure in Experiment 1, with the following notable exceptions. First, singular nonwords ending in non-alternating intermediate sounds (voiced stops in the Stops sub-group, voiceless fricatives in the Fricatives sub-group), and the corresponding sounds in the Control condition, were included in the exposure and verification phases. These singular nonwords were paired with non-changing plural options during the exposure phase. Participants were still required to reach 80% accuracy in the verification phase. Including intermediate sounds in the verification phase likely made the task harder, a point I will return to in section 3.5.3 below. Finally, the generalization phase contained only 64 trials in Experiment 2 as opposed to 72 trials in Experiment 1.

74

3.5.2 Results

Only results from the generalization phase were analyzed. Analyses were conducted using mixed-effects logistic regression models as in Experiment 1.

3.5.2.1 Trained sounds – alternating and filler

Performance on these sounds was not the focus of this study and, in fact, was partially used as an exclusion criterion. Recall that four participants were excluded from the Saltatory condition even though they reached the 80% criterion in the verification phase because their accuracy on the trained alternations in the generalization phase was below 10% (see section 3.5.1.1). Among the remaining participants, accuracy on the trained alternations and fillers sounds was similar in the Saltatory condition (94.9%) and in the Control condition (95.2%).

3.5.2.2 Trained sounds – non-alternating intermediate

Recall that the primary purpose of Experiment 2 was to determine whether participants would find it difficult to learn that intermediate sounds did not alternate when provided with explicit evidence during training. If so, participants were predicted to incorrectly choose the changing plural option more often for the intermediate sounds in the Saltatory condition than for the comparable sounds in the Control condition. Figure 6 shows how often participants chose the changing plural option (in this case, an incorrect response) for trained non-alternating intermediate sounds (Saltatory condition) and comparable control sounds (Control condition) in Experiment 2, sorted by Exposure Group (Stops sub-group and Fricatives sub-group). Overall, we see that participants in the Saltatory condition made more errors than participants in the

75

Control condition as predicted, both for the Stops sub-group (20.8% vs. 6.7% errors) and the Fricatives sub-group (38.8% vs. 18.3% errors).

To evaluate these differences, a mixed logit model was fitted to predict the log odds of an error (i.e., a changing plural response). The final model included a fixed effect for Condition (Saltatory vs. Control) and a fixed effect for Exposure Group (Stops sub-group vs. Fricatives sub-group). The Condition x Exposure Group interaction was not included because it was not a significant predictor ($p = .76$), and a likelihood ratio test indicated that it did not significantly improve the fit of the model when compared to the model without an interaction effect, $\chi^2(1) = .09$, $p = .76$. Random intercepts for subjects and for items were included in the model because they significantly improved model fit ($\chi^2(1) = 20.68$, $p < .001$ for subjects, $\chi^2(1) = 128.01$ $p < .001$ for items). By-subject random slopes were not included in the model because the comparisons were fully between-subjects.

The fixed effects for the final model are provided in Table 3. The significant negative intercept reflects the fact that participants had an overall low rate of errors in the Control condition (which acts as the baseline in this model). Condition was a significant predictor in the model, indicating that participants made more errors (by incorrectly changing intermediate sounds) in the Saltatory condition compared to the Control condition, as predicted. Exposure Group was also a significant predictor, reflecting the fact that overall, participants in the Fricatives sub-group made more errors than participants in the Stops sub-group. The lack of a significant Condition x Exposure Group interaction effect indicates that the difference in accuracy between the Saltatory condition and the Control condition holds for both the Stops sub-group and the Fricatives sub-groups. Indeed, Condition remains a significant predictor in models

76

run on subsets of the data containing only the Stops sub-group (Estimate = 1.75, z = 2.20, p = .03) or only the Fricatives sub-group (Estimate = 1.35, z = 2.56, p = .01).

Overall, the results show that even with explicit evidence, participants found it harder to learn that sounds did not alternate if the sounds were intermediate between alternating sounds (Saltatory condition) than if they were not intermediate (Control group). This main effect was true regardless of whether the intermediate sounds were voiced stops (Stops sub-group) or voiceless fricatives (Fricatives sub-group).

Figure 6. Percent of trials in which the changing plural option was chosen for trained intermediate sounds (Saltatory condition) and comparable non-intermediate sounds (Control condition) in Experiment 2 by sub-group. Individual results (diamonds) and overall means (bars) are provided.

Table 3. Summary of the fixed effects in the final model for trained intermediate sounds (and control sounds) in Experiment 2.

Predictor	Estimate	Standard error	z-value	p-value
Intercept	−3.66	.45	−8.21	<.001
Condition = *Saltatory*	1.49	.45	3.31	.001
Exposure group = *Fricatives sub-group*	1.57	.51	3.09	.002

3.5.2.3 Effect of amount of exposure

As in Experiment 1, participants in Experiment 2 received varying amounts of exposure due to the nature of the design. To address the possibility that the amount of exposure affected participants' accuracy on intermediate sounds, the logit model in Table 4 was rerun with an added fixed effect for number of cycles through the exposure phase. Amount of exposure was not a significant predictor in the model ($p = .20$), and including the factor in the model did not significantly improve model fit, $\chi^2(1) = 1.58$, $p = .21$. Including amount of exposure along with its interactions with the other fixed effects also failed to significantly improve model fit, $\chi^2(3) = 4.60$, $p = .20$, and none of the new factors reached significance in the model. Thus, we may conclude that amount of exposure did not have a significant effect on how often participants changed intermediate sounds (and comparable sounds in the Control condition) in Experiment 2.

3.5.2.4 Untrained sounds

A secondary goal for Experiment 2 was to replicate the results for untrained sounds found in Experiment 1. Like in Experiment 1, untrained sounds in Experiment 2 were intermediate between alternating sounds in the Saltatory condition, but were not intermediate in the Control condition, so the pattern of results for untrained sounds should be similar to the one found in Experiment 1. Recall that unlike in Experiment 1, however, each participant in Experiment 2 only had one type of untrained sound, either stops or fricatives, whereas participants in

78

Experiment 1 had both types. Figure 7 presents the percent of trials that participants chose the changing plural option for words ending in untrained sounds in Experiment 2. Overall, we see that the values in Experiment 2 are similar to those in Experiment 1 (see Figure 4).

A mixed logit model was run to check for potential differences in how participants treated untrained sounds in Experiment 1 and Experiment 2. The random effect structure included intercepts for subjects and by-subject slopes for Sound Type. The by-subject slopes were included because they significantly improved model fit, $\chi^2(2) = 71.83$ $p < .001$. Adding random intercepts for items did not significantly improve model fit, $\chi^2(1) = 3.74$ $p = .053$.

The fixed effects of the final model are summarized in Table 4. The final model included a fixed effect for Condition (Potentially Saltatory/Saltatory vs. Control), a fixed effect for Sound Type (untrained stops vs. untrained fricatives), and a Condition x Sound Type interaction. The significant negative intercept indicates that participants in the Control condition (which acts as the baseline for this model) changed untrained sounds at a low rate. Condition is a significant predictor in the model, indicating that participants in the Potentially Saltatory/Saltatory condition changed untrained sounds more frequently than those in the Control condition. The significant interaction effect indicates that untrained stops were changed more often than untrained fricatives, but only in the Potentially Saltatory/Saltatory conditions. This model looks very similar to the one conducted for Experiment 1 (Table 2). Indeed, neither the factor for Experiment (Exp. 1 vs. Exp. 2) nor its associated interaction effects were significant predictors in the model, and including them did not significantly improve model fit. Overall, the results for untrained sounds in Experiment 2 replicated the basic findings from Experiment 1.

Figure 7. Percent of trials in which the changing plural option was chosen for untrained target sounds according to Condition (Saltatory or Control) and sub-group in Experiment 2. Individual results (diamonds) and overall means (bars) are provided.
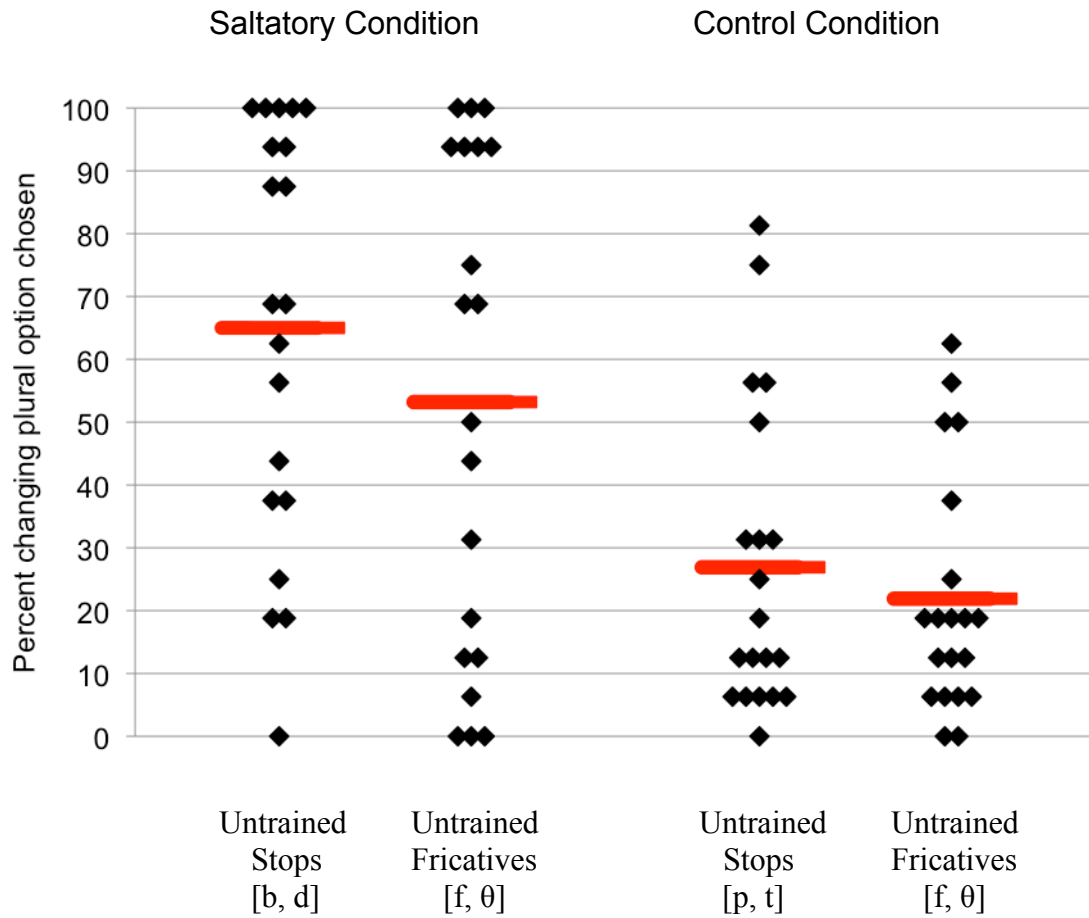
Table 4. Summary of the fixed effects in the final model for untrained sounds in Experiments 1 and 2 combined.

Predictor	Estimate	Standard error	z-value	p-value
Intercept	−2.11	.36	−5.87	<.001
Condition = *Potentially Saltatory/Saltatory*	1.93	.50	3.87	<.001
Sound Type = *Untrained stops*	.22	.46	.47	.64
Interaction = *Potentially Saltatory/Saltatory & Untrained stops*	1.50	.64	2.34	.02

3.5.3 Discussion

Experiment 2 accomplished two main objectives. First, it replicated the primary finding of Experiment 1: untrained sounds were changed more frequently when they were intermediate between alternating sounds (Saltatory condition) than when they were not intermediate (Control condition). Second, Experiment 2 expanded on the findings from Experiment 1 by showing that the desire to change intermediate sounds is strong enough to have an effect even when there is evidence in the input that intermediate sounds do not change. Thus, the saltation avoidance effect not only affects how learners interpret ambiguous input, but it also affects how well learners acquire patterns provided explicitly in the input.

Two points related to the experimental design speak further to the robustness of the saltation avoidance effect observed in Experiment 2. First, participants had to reach 80% accuracy in the verification phase even though it included the trained intermediate sounds. To advance to the generalization phase, participants had to respond correctly for at least some of the trials with intermediate sounds in the verification phase. Therefore, it is striking that participants still made a significant number of errors on intermediate sounds once they reached the generalization phase.

Relatedly, participants who failed to reach the criterion in the verification phase within 50 minutes did not complete the experiment, and including the intermediate sounds in the verification phase increased the overall difficulty of the task. As a result, the attrition rate in Experiment 2 was fairly high. The attrition has the effect of potentially limiting the representativeness of the sample, but in fact, it does so in a way that is desirable. Those participants who reached the generalization phase (and thus were used in the analysis) consisted of the learners who were best able to learn the new patterns. Yet even among the best learners,

there were significantly more errors on the intermediate sounds (Saltatory condition) than on the non-intermediate sounds (Control condition). If anything, excluding intermediate sounds from the verification phase or relaxing the 80% criterion would likely increase the difference in accuracy observed between the Saltatory and Control condition. It is quite possible that several participants failed to complete the experiment because they were unable to learn that intermediate sounds did not alternate, and if so, including them would only enhance the difference. The fact that a significant difference was found using the more restrictive inclusion criterion in the current study implies that the effect is quite robust.

One aspect of the results was unexpected—namely, there were more errors on intermediate fricatives than on intermediate stops. Recall that in Experiment 1, the opposite direction was found: participants changed intermediate stops more often than intermediate fricatives. In the case of Experiment 2, the higher amount of errors for intermediate fricatives is likely due to English orthography. During the post-experiment debriefing, the vast majority of participants reported their strategy in terms of letters, that is, they would say things like "*p* becomes *v*" rather than making the [p] and [v] sounds. The dental fricatives [θ] and [ð] are both written as <th> in English, which plausibly led to confusion when participants were forced to remember generalizations involving these sounds. The generalization "*th* becomes *th*" is ambiguous between four possible mappings: [θ] to [ð], [θ] to [θ], [ð] to [θ], and [ð] to [ð]. At test, participants may have experienced difficulty recalling which mapping was the one that they had actually learned. In Experiment 1, the orthographic ambiguity of [θ] and [ð] was not an issue. In that experiment, [θ] was always an untrained sound, meaning that participants never needed to remember a generalization involving [θ]. At test, there would be less of a need to consider the orthography of [θ] and [ð] because they could just compare the two sounds directly.

Support for this explanation comes from the fact that the percentage of errors for [θ] far exceeds the percentage of errors for [f] in the Fricatives sub-group of Experiment 2 (in the Saltatory condition, [θ]: 49% changed vs. [f]: 28% changed; in the Control condition, [θ]: 26% changed vs. [f]: 11% changed). This large discrepancy between [θ] and [f] was not found in Experiment 1 (Potentially Saltatory condition, [θ]: 41% vs. [f]: 49%; Control condition, [θ] 18% vs. [f]: 14%), suggesting that it was due to greater confusion with the dental sounds when participants had to actually learn generalizations involving them during training. Crucially, the saltation avoidance effect (i.e., more errors in the Saltatory condition than in the Control condition) still occurred in the Fricatives sub-group, independent of the greater number of errors due to orthographic ambiguity.

Finally, there are two logically possible ways that participants could exhibit difficulty learning saltatory alternations: (1) they could wrongly change intermediate sounds, or (2) they could fail to properly change the alternating sounds. This experiment focused on the first case— how participants treated intermediate sounds. By ensuring that participants learned the alternations in question, the design did not permit a systematic investigation of how difficult it is to learn the alternation itself, but follow-up experiments could be designed to do so. There is indirect evidence from the current experiments supporting the possibility that the alternations themselves are harder to learn. First, the attrition rate (i.e., the number of participants never making criterion in the verification phase) in the Saltatory condition of Experiment 2 was far greater (n = 13 out of 33) than the attrition rate in the Potentially Saltatory condition of Experiment 1 (n = 2 out of 22), implying that the task was much harder with the addition of the non-alternating, intermediate sounds. Part of this difference is likely due to errors on the

intermediate sounds themselves, as mentioned, but some of it may also be due to the increased difficulty of learning the alternating sounds.

Second, even among those participants who made it into the generalization phase in Experiment 2, four participants had to be excluded from the Saltatory condition because they apparently "forgot" the alternations that they learned during exposure (i.e., they had less than 10% accuracy on at least one of the alternating sounds in the generalization phase, see section 3.5.1.1). These observations suggest that the difficulty associated with learning saltatory alternations may not be limited to intermediate sounds (as demonstrated in this study), but may also be reflected in how quickly learners acquire the alternations themselves. A modified version of the current experiment, or one designed to measure speed of acquisition (e.g., Skoruppa et al., 2011), would be useful to corroborate this prediction, but I leave such an investigation for future research.

3.6 General discussion

The purpose of this study was to determine experimentally whether adults have a learning bias that disfavors saltatory phonological alternations. The most striking aspect of the results was that participants extended learned alternations to untrained intermediate sounds (but not to comparable non-intermediate sounds) without evidence in the input (Exp. 1), and in some cases *contrary* to evidence in the input (Exp. 2). To summarize, in Experiment 1 learners changed untrained sounds much more frequently when doing so would avoid a saltatory alternation (Potentially Saltatory condition) than when there was no chance of a saltatory alternation (Control condition). Experiment 2 showed that even with explicit training, participants had greater difficulty learning that sounds did not change if they were intermediate between two

alternating sounds (Saltatory condition) than if they were not intermediate (Control condition). Taken together, the results from these experiments provide strong evidence that people learn novel alternations with a preference for avoiding saltation. This study adds to a growing body of literature showing that phonological learning is constrained by biases (Saffran & Thiessen, 2003; Wilson, 2006; Zuraw, 2007; Finley, 2008; Finley & Badecker, 2008; Moreton, 2008; Hayes et al., 2009; Skoruppa, Lambrechts, & Peperkamp, 2011; Skoruppa & Peperkamp, 2011; Baer-Henney & van de Vijver, 2012; Hayes & White, 2013).

Given what we know about saltatory alternations, our theory of phonological learning must be able to account for two facts: (1) saltatory alternations are attested in natural languages (see section 2.2) and therefore must be learnable, and (2) they are dispreferred during learning, as demonstrated in this study. From the results of this study, it is clear that participants were biased against saltation. But which type of formal bias is responsible for this saltation avoidance? Put another way, how should we account for the saltation avoidance in formal learning models?

One way to deal with the problem of saltatory alternations is by implementing a hard bias, that is, by ruling them out altogether, as in Peperkamp, Le Calvez, Nadal, and Dupoux (2006a). They implemented a computational model that learned which sounds were context-dependent variants of other sounds by looking for complementary distributions. Two sounds were considered to be allophonic variants if they rarely occurred in the same phonological environment, that is, if their distributions in the input had little or no overlap. To prevent the model from learning spurious pairings (e.g., between [h] and [ŋ] in English, as discussed in the section 3.1), Peperkamp et al. equipped the basic statistical model with two linguistic filters. One of these filters prohibited mappings between two sounds if an intermediate sound existed between them, where an intermediate sound was defined similarly to the way it is defined here.

The filter, which effectively banned saltatory alternations, improved the model's performance because it was successful at excluding spurious mappings. However, it would also prevent the model from learning that [t] and [ð] alternate in Campidanian Sardinian due to the presence of intermediate [d]. That conclusion would, of course, be incorrect in the case of Campidanian Sardinian. Given the existence of languages with saltatory alternations, it must be possible for a child to learn them; thus, an absolute ban on saltatory alternations, like the one proposed by Peperkamp et al., is not tenable.

To account for the fact that saltatory alternations are dispreferred during learning but not unlearnable, we most likely need a soft bias. The idea that soft biases have a role in phonological learning has been growing in the literature (e.g., see Wilson, 2006; Zuraw, 2007; Finley & Badecker, 2008; Moreton, 2008; Hayes et al., 2009; Baer-Henney & van de Vijver, 2012; Hayes & White, 2013). This literature is reviewed in more detail in section 1.2. I argue that the appropriate explanation for saltation avoidance is a substantive bias based on the principle of minimal modification in Steriade's P-map theory (2001/2008). In Chapter 4, I implement this bias in a MaxEnt learning model; here I will provide brief sketch of the overall concept.

3.6.1 Substantive bias as an explanation for saltation avoidance

It has long been noted that alternating sounds tend to be highly similar, and alternations between dissimilar sounds are less common than those between similar sounds (e.g., Trubetzkoy, 1939/1969). Recall that a saltatory alternation is a particularly striking counterexample to the principle of minimal modification, proposed by Steriade (2001/2008) as an explanation for why alternations between dissimilar sounds are uncommon. Skoruppa, Lambrechts, and Peperkamp

86

(2011) provided experimental evidence that language learners are sensitive to the similarity between sounds when learning novel alternations in an artificial language. In their study, adults were able to learn novel alternations between sounds differing in a single feature (e.g., [p ~ t]) more easily than sounds differing in two or more features (e.g., [p ~ z]). These results are straightforwardly predicted by a substantive bias based on the principle of minimal modification expressed in the P-map (Steriade, 2001/2008). If learners are biased to prefer alternations between similar sounds, they should find [p ~ t] easier to learn than [p ~ z].

The current study did not focus on the similarity of the alternating sounds directly, but rather on how learners treated sounds that were intermediate between alternating sounds. Still, the same mechanism could plausibly explain why participants learning [p ~ v] were biased to assume [b ~ v]. Conceptually, the similarity account works as follows. By learning that voiceless stops [p, t] changed to voiced fricatives [v, ð] between vowels, participants had evidence for an active restriction on stops between vowels. When faced with voiced stops [b, d] at test, participants noticed that they also violated the restriction on stops between vowels; moreover, voiced stops are more similar to voiced fricatives than are voiceless stops, resulting in a tendency to change the voiced stops. Intuitively, if a phonological constraint (e.g, no stops between vowels) warrants alternations between dissimilar sounds, then it should also warrant alternations between more similar sounds. The same logic would apply to the voiceless fricatives [f, θ], except the relevant phonological constraint would be a ban on voiceless sounds between vowels. In the Control condition, extending the alternation to untrained sounds is correctly predicted to occur less frequently because the same logic does not hold in the reverse direction: evidence for an alternation between highly similar sounds does not necessarily license an alternation between less similar sounds.

3.6.2 Anti-complexity bias

Another possible bias that could be relevant to the results of this study is an anti-complexity bias. There is extensive evidence that humans prefer simple solutions to more complex solutions within a number of cognitive domains, and indeed, the results of many artificial grammar studies showing differences in phonological learning can potentially be attributed to a preference for simpler patterns. A general overview of complexity bias is provided in section 1.2.2.3. For an extensive review, see Moreton & Pater (2012a, 2012b).

In the current study, neither of the alternations in the Potentially Saltatory condition or the Control condition in Experiment 1 is more complex than the other in terms of the classes of sounds being targeted (and likewise in Experiment 2). In each condition, one pair of segments alternates with one other pair of segments in a parallel way: [p, t] → [v, ð] vs. [b, d] → [v, ð]. However, if learners track the number of phonological features that change, rather than just the ones needed to target the class of sounds undergoing the change, the avoidance of saltatory alternations seen in Experiment 1 might be viewed as an effect of complexity, rather than an effect of phonetic similarity *per se*.[20]

Under the feature-counting analysis, participants in Experiment 1 were considering (at least) two possible rules in each of the conditions: a general rule and a narrow rule. In the Potentially Saltatory condition, those rules may be formalized as follows:

---

[20] Thanks to Elliott Moreton for pointing out this possible analysis to me.

(26) *General and narrow rules for the Potentially Saltatory condition of Experiment 1*

a. General rule (fewer features):

$$[\text{–sonorant}] \longrightarrow \begin{bmatrix} +\text{continuant} \\ +\text{voice} \end{bmatrix} \Big/ \text{V} \underline{\quad} \text{V}$$

Effect (for labials): "{p, f, b} become [v] between vowels."

b. Narrow rule (more features):

$$\begin{bmatrix} -\text{sonorant} \\ -\text{voice} \end{bmatrix} \longrightarrow \begin{bmatrix} +\text{continuant} \\ +\text{voice} \end{bmatrix} \Big/ \text{V} \underline{\quad} \text{V}$$

Effect (for labials): "{p} becomes [v] between vowels."

The narrow rule (26b) requires one extra feature compared to the general rule (26a). Thus, participants may have preferred the general rule, leading them to change intermediate sounds.

In the Control condition, participants also had the option of a general rule (27a) and a narrow rule (27b), but in this case the two rules involve an equal number of total features. If participants prefer rules with fewer total features, they would have no preference between these rules, even though one applies to a general class of sounds and the other applies to only a single sound. This analysis crucially depends on the ability to track features that are changing in addition to features used to classify the targeted group of sounds; otherwise, no differences would be expected between the two conditions.

(27) *General and narrow rules for the Control condition of Experiment 1*

a. General rule:

$$[-\text{sonorant}] \longrightarrow \begin{bmatrix} +\text{continuant} \\ +\text{voice} \end{bmatrix} \Big/ \quad V \underline{\quad} V$$

Effect (for labials): "{p, f, b} become [v] between vowels."

b. Narrow rule:

$$\begin{bmatrix} -\text{sonorant} \\ +\text{voice} \end{bmatrix} \longrightarrow [+\text{continuant}] \Big/ \quad V \underline{\quad} V$$

Effect (for labials): "{b} becomes [v] between vowels."

However, even if the saltation avoidance effect could be explained by an anti-complexity bias, it is unclear how the anti-complexity account would explain the preference observed in Experiment 1 for changing voiced stops (to voiced fricatives) more often than changing voiceless fricatives. The feature-counting approach predicts no difference preference for changing one type of sound over the other (or perhaps the opposite difference, see fn. 17) because both differ from the target (voiced fricatives) by an equal number of features. Accounting for this difference would require appealing to phonetic and/or perceptual similarity, implying that we would still need to have a substantive bias in the model.

3.6.3 Task considerations

Some remaining issues related to the task used in these experiments warrant discussion. First, these studies used an explicit two-alternative forced-choice task to test participants' learning. The advantage of this task is that the data analysis is straightforward, but it has two

potential downsides: (a) participants may wish to provide answers that are not among the response options, and (b) participants may not have considered one or both of the response alternatives if they had not been provided. In section 4.5, I expand on these experiments by replicating Experiment 1 in a production task, where participants are free to offer any response.

Second, the results from these experiments provide strong evidence that <u>adults</u> have a bias against saltatory alternations, but ultimately we are most interested in how children acquire language. Like any artificial language study with adult participants, this work faces the limitation that adults may use strategies in the experimental task that are not available to children during acquisition of a first language. In particular, adults may bring native language knowledge to bear on the task or they may use non-linguistic problem solving strategies. It is worth noting that English does not have the alternations tested here, with the exception of the marginally productive fricative voicing rules for plurals (e.g., *half* [hæf] ~ *halves* [hævz]; Becker et al., 2012) and noun-verb pairs (e.g., *teeth* [tiθ] ~ *teethe* [tið]). Despite this marginal evidence for a fricative voicing alternation in English, however, participants in Experiment 1 actually had a preference for spirantizing stops compared to voicing fricatives.

Still, we cannot be sure that participants were not bringing native language knowledge to the task in some form. There is also no way to know for sure the extent to which participants were employing linguistic mechanisms rather than using non-linguistic problem solving strategies (although we would still be left explaining why these strategies resulted in saltation avoidance). For these reasons, it is important to test for saltation avoidance in infants who are just beginning to learn alternations. This task is undertaken in Chapter 5.

3.7 Chapter summary

This study has provided experimental evidence that adults are biased against learning saltatory alternations when learning an artificial language, adding to the growing body of literature showing that language learners display biases against certain phonological patterns. Given these results, models of phonological learning must be able to account for the dispreferred status, yet ultimate learnability, of saltatory alternations. Augmenting models of phonological learning with a substantive bias based on the principle of minimal modification, that is, one that assigns greater prior likelihoods to alternations between sounds with greater perceptual similarity, appears promising as a way to account for the facts observed in this study. This approach is taken in the next chapter.

3.8 Appendix

This section provides representative sample stimuli from Experiment 1. A full list of stimuli for Experiments 1 and 2 is available at the author's website, currently at:

http://www.linguistics.ucla.edu/people/grads/jwhite/papers.htm

Exposure phase

Potentially Saltatory condition			Control condition		
Type	*Examples*		*Type*	*Examples*	
	Singular	*Plural*		*Singular*	*Plural*
18 p → v	*kamap*	*kamavi*	*18 b → v*	*kamab*	*kamavi*
	nisup	*nisuvi*		*nisub*	*nisuvi*
18 t → ð	*ɹamit*	*ɹamiði*	*18 d → ð*	*ɹamid*	*ɹamiði*
	kunit	*kuniði*		*kunid*	*kuniði*
36 fillers (same in both conditions)	*luman*	*lumani*			
	gunam	*gunami*			
	misil	*misili*			
	kaluɹ	*kaluɹi*			
	kuɹas	*kuɹasi*			
	ɹamiʃ	*ɹamiʃi*			

Verification phase (subset from exposure phase, correct answer in bold font)

Potentially Saltatory condition				Control condition			
Type	Examples			Type	Examples		
	Singular	Non-changing plural option	Changing plural option		Singular	Non-changing plural option	Changing plural option
8 p-final	kamap	kamapi	**kamavi**	8 b-final	kamab	kamabi	**kamavi**
	nisup	nisupi	**nisuvi**		nisub	nisubi	**nisuvi**
8 t-final	ɹamit	ɹamiti	**ɹamiði**	8 d-final	ɹamid	ɹamidi	**ɹamiði**
	kunit	kuniti	**kuniði**		kunid	kunidi	**kuniði**
16 fillers (same in both conditions)	luman	**lumani**	lumaði				
	gunam	**gunami**	gunavi				
	gimal	**gimali**	gimaði				
	kaluɹ	**kaluɹi**	kaluvi				
	kuɹas	**kuɹasi**	kuɹaði				
	ʧimaɹ	**ɹamiʃi**	ɹamivi				

Generalization phase

Potentially Saltatory condition				Control condition			
Type	Examples			Type	Examples		
	Singular	Non-changing plural option	Changing plural option		Singular	Non-changing plural option	Changing plural option
12 p-final	sulap	sulapi	sulavi	12 b-final	sulab	sulabi	sulavi
	kiʃap	kiʃapi	kiʃavi		kiʃab	kiʃabi	kiʃavi
12 t-final	gumut	gumuti	gumuði	12 d-final	gumud	gumudi	gumuði
	ʃaɹut	ʃaɹuti	ʃaɹuði		ʃaɹud	ʃaɹudi	ʃaɹuði
6 b-final	talab	talabi	talavi	6 p-final	talap	talapi	talavi
6 d-final	masid	masidi	masiði	6 t-final	masip	masiti	masiði
6 f-final	tunuf	tunufi	tunuvi	6 f-final	tunuf	tunufi	tunuvi
6 θ-final	paɹuθ	paɹuθi	paɹuði	6 θ-final	paɹuθ	paɹuθi	paɹuði
24 fillers (same in both conditions)	niʃin	niʃini	niʃiði				
	tasam	tasami	tasavi				
	baɹul	baɹuli	baɹuði				
	funiɹ	funiɹi	funivi				
	ɹanus	ɹanusi	ɹanuði				
	θanaʃ	θanaʃi	θanavi				

Accounting for saltation within phonological theory:
A MaxEnt learning model with a substantive bias based on the P-map

4.1 The theoretical challenge posed by saltation

The case of saltation poses a challenge for current phonological frameworks. The first problem is having a theory that is able to generate saltations in the first place. In Chapter 2, we saw cases where saltations are attested in natural languages. In order for saltations to be attested, it must be possible for learners to acquire a saltatory system. Thus, our theory must be able to accommodate grammars that allow phonological saltation.

Second, in Chapter 3 we saw results from artificial language experiments suggesting that saltatory alternations are dispreferred and difficult to learn. The goal of phonological theory is not only to explain which patterns are possible or impossible in languages, but also to explain why certain patterns, while possible, may be difficult for learners or otherwise dispreferred or rare in the world's languages (e.g., see Kiparsky, 1982, p. 59–60). Therefore, we need a theory that (a) predicts that saltation is a possible (i.e., learnable) phonological pattern, and (b) explains why saltation, though learnable, is dispreferred by the learner.

In this chapter, I argue for a theoretical framework with three components. First, traditional feature-based faithfulness constraints are augmented with a more powerful set of *MAP constraints (Zuraw, 2007), which are capable of penalizing correspondences between any two individual sounds. These constraints are necessary to allow saltation in the first place. Second, I adopt a substantive bias based on Steriade's P-map (2001/2008), which makes phonological

changes between perceptually similar sounds easier to learn than those between dissimilar sounds. Third, I use a Maximum Entropy (MaxEnt) learning model as the grammatical framework. The MaxEnt model allows a soft bias to be implemented by way of the model's prior (as in Wilson, 2006); it also allows us to test the model's predictions through learning simulations. Overall, I will show that a model with these components is successful at predicting the desired learning behavior: saltations are initially dispreferred, but with enough data, they are ultimately learnable.

I will first provide an overview of MaxEnt models in general and then outline the basics of the model's architecture as I have implemented it. I will test the model's predictions by feeding it the same training data as the participants received in the artificial language experiments from Chapter 3. Then, I will provide a more difficult test of the model by presenting results from a production experiment. Finally, I will consider further implications of the model.

4.2 The MaxEnt model

4.2.1 Overview of MaxEnt grammar models

Maximum entropy models (also known as log-linear models) describe a general type of statistical model that has been used in a wide range of areas. They were first used to model phonological grammars by Goldwater and Johnson (2003) and have since been used in several other studies (e.g., Wilson, 2006; Hayes & Wilson, 2008; Hayes et al., 2009; Martin, 2011; Hayes, Wilson, & Shisko, 2012; Pater et al., 2012).

For each input, the MaxEnt model generates a probability distribution over the set of output candidates based on their violations of a set of weighted constraints. Specifically, for some input x, it assigns a probability to each output candidate y as follows:

$$(28) \quad \Pr(y \mid x) = \frac{\exp\left(-\sum_{i=1}^{m} w_i C_i(y, x)\right)}{Z} \quad , \text{ where}$$

$$Z = \sum_{y \in Y(x)} \exp\left(-\sum_{i=1}^{m} w_i C_i(y, x)\right)$$

Based on the formula in (28), the method for calculating the probability of an output candidate y for an input x can thus be described as follows. First, for each constraint, multiply the weight w_i of that constraint by the number of times the input/output pair violates the constraint, $C_i(y, x)$, and then sum over all constraints $C_1...C_m$. This summed value, $\sum w_i C_i(y, x)$, is comparable to the Harmony value from Harmonic Grammar (Legendre et al., 1990; Smolensky & Legendre, 2006; Pater, 2009b) and has also been called a Penalty score (Hayes & Wilson, 2008). Raise e to the negative Penalty score and finally divide the result by the sum over all possible output candidates (all y in the set $Y(x)$) for that input x. The sum over all output candidates is typically represented as Z.

As implemented here, this framework has a clear connection to Optimality Theory (OT; Prince & Smolensky, 1993/2004), as has been discussed by others (e.g., Eisner, 2000; Johnson, 2002; Goldwater & Johnson, 2003; Hayes et al., 2009). As such, it is sometimes referred to as MaxEnt-OT. The model is assumed to have a component comparable to GEN (i.e., $Y(x)$ in the formula in (28) above), which generates the set of output candidates for a given input form. The set of candidates is then evaluated on the basis of the grammar.

In classical OT, candidates are evaluated based on a strict ranking of the constraints, such that one candidate is judged the winner if it is preferred by (i.e., has fewer violations of) the highest ranked constraint in the hierarchy (Prince & Smolensky, 1993/2004). Constraints lower

in the hierarchy have an influence on the outcome only if all higher ranked constraints have no preference among the candidates (i.e., only if all candidates have the same number of violations for each of the higher ranked constraints). Only one candidate (or set of candidates with identical violation profiles) in classical OT is declared the winner; all other candidates are losers. Thus, classical OT is not an effective framework for modeling variation in which a single input may have multiple possible outputs.

The EVAL component of the MaxEnt model generates a probability distribution over all possible candidates for a given input, and unlike in classical OT, the total probability may be divided unequally across different candidates. If the constraint weights are sufficiently different from one another[21], the MaxEnt grammar will mimic the strict constraint rankings from classical OT, such that only one effective winner will emerge with a probability very close to 1.[22] In fact, it is possible to generate a MaxEnt simulation for any categorical outcome analyzed with classical OT as long as there is a finite limit on the number of constraint violations (Johnson, 2002). However, if the constraint weights are similar to each other, then multiple candidates will be assigned probabilities that are not vanishingly small. In such cases, the model predicts that there will be variation in which output will be chosen; moreover, the predicted probabilities for each output candidate can be compared to real data collected from a corpus or an experiment, such as the experiments reported in Chapter 3.

MaxEnt is one of several frameworks that have been proposed to account for variable outputs in phonology. In particular, several other modifications to classical OT have been

---

[21] To achieve this, constraint weights need to be spaced at roughly exponential increments, see Johnson, 2002; Goldwater and Johnson, 2003.

[22] The probability can never actually reach 1 because other candidates must receive some probability, even if vanishingly small (i.e., the numerator in (28) can never reach 0 for any given candidate).

proposed to handle variation, such as strata with freely ranked constraints (Anttila, 1997), floating constraints (Ross, 1996; Nagy & Reynolds, 1997), constraints with strictness bands (Hayes & MacEachern, 1998; Hayes, 2000), and Stochastic OT with its associated Gradual Learning Algorithm (GLA; Boersma, 1997; Boersma & Hayes, 2001). MaxEnt models, however, have certain characteristics that differentiate them from other approaches to modeling variation. First, MaxEnt models involve summing the violations of multiple weighted constraints (as in Harmonic Grammar; Legendre, et al., 1990) rather than following a strict ranking hierarchy. As a result, MaxEnt models have the property of cumulative constraint interaction, often called "ganging," whereby multiple violations of lower constraints can add up to overcome a violation of a constraint with a greater weight (Hayes & Wilson, 2008; Pater, 2009b; see also Flemming, 2012, for an argument that the ganging property is not desirable). Second, MaxEnt models are particularly attractive because they are associated with a learning algorithm (Berger et al., 1996) that provably converges on the objectively "best" grammar, which in MaxEnt is defined as the set of constraint weights that maximize the probability of the training data (taking into account the prior). By comparison, it has been shown that the GLA sometimes fails to converge on a grammar, even when a grammar exists that could, in principle, account for the data (e.g., Pater, 2008).[23]

4.2.2 Learning the weights

Given a set of constraints and the observed data, the learning problem for the MaxEnt model is to find the weights that maximize the probability of the observed data (thereby minimizing the

---

[23] A modification to the GLA by Magri (2012) allows it to successfully handle the case put forth by Pater (2008). However, in unpublished work, Bruce Hayes (personal communication) has discovered another case in which the GLA fails to converge, even when Magri's modification is used.

probability of unobserved data). The probability of the observed data, D, is calculated by taking the product of the model-predicted conditional probabilities of each output observed during training given its input, $\{(y_1 \mid x_1) \ldots (y_n \mid x_n)\}$:

(29) $\Pr(D) = \displaystyle\prod_{j=1}^{n} \Pr(y_j \mid x_j)$

This calculation is computed on the basis of observed tokens, so 100 examples of the input/output pair (b | p) during training will have a greater effect on the model than only one example. Because probabilities are being multiplied, the $\Pr(D)$ calculated in (29) is always extremely small, so in practice the calculation is done by taking the sum of the log probabilities of each output given its input:

(30) $\log \Pr(D) = \displaystyle\sum_{j=1}^{n} \log \Pr(y_j \mid x_j)$ $\qquad$ (equivalent to log (29))

The model also takes into account a regularizing bias term, often called a "prior," during learning. The prior term is a Gaussian distribution over each constraint weight, defined in terms of a mean μ and a standard deviation σ:

(31) $\displaystyle\sum_{i=1}^{m} \frac{(w_i - \mu_i)^2}{2\sigma_i^2}$

The result of (31) is subtracted from the probability calculated in (30). The μ for each constraint acts as its *a priori* preferred weight, which is subtracted from the constraint's learned weight, w; the difference in actual and preferred weight is then squared. Thus, as constraints vary more from their μ, the penalty imposed by the prior increases. The value of σ^2 determines how tightly each

constraint's weight is constrained to its μ. Because it is in the denominator, lower values of σ^2 result in a greater penalty for weights that vary from their μ. As a result, low values of σ^2 mean that more data are required to move the weights away from μ during learning. Higher values of σ^2 mean that the weights have more freedom to vary from their μ. Overall, the prior acts as a penalty that increases as constraint weights diverge from their *a priori* preferred weights.

When the prior is uniform across all constraints (and σ^2 is not set very high), the model prefers grammars in which weight is distributed among each of the constraints and ample amounts of data are needed for constraints to reach relatively extreme weights. For this reason, Gaussian priors are commonly used in MaxEnt models as a way to prevent overfitting. (discussed, e.g., by Goldwater & Johnson, 2003). In my model, constraints may each receive a different μ, so the prior also serves as a means of implementing a substantive learning bias (see section 4.3 below), following previous work by Wilson (2006).

With the inclusion of a Gaussian prior, the goal of learning then is to choose the set of constraint weights that maximize the objective function in (32), in which the prior term in (31) is subtracted from the log probability of the observed data (the function in (30)):

(32) $\quad [\sum_{j=1}^{n} \log \Pr (y_j \mid x_j)] \; - \; [\sum_{i=1}^{m} \frac{(w_i - \mu_i)^2}{2\sigma_i^2}]$

The search space of log likelihoods is provably convex, meaning that there is always one objective set of weights that will maximize the function in (32), and this set of weights can be found using any standard optimization strategy (Berger et al., 1996). To implement the model, I

used the MaxEnt Grammar Tool[24], which uses the Conjugate Gradient algorithm (Press et al., 1992) to find the weights during learning.

4.2.3 *M_AP Constraints

The real problem underlying classical OT's inability to <u>generate</u> saltation is the set of traditional feature-based faithfulness constraints, such as IDENT(cont) and IDENT(voice). In fact, the same problem persists in theories that abandon strict dominance in favor of weighted constraints, such as Harmonic Grammar (section 2.4.3). The reason for this becomes clear by considering a simple example. Consider the saltation in which /p/ surfaces as [β], but /b/ remains [b]. Assuming traditional IDENT constraints, the faithfulness violations amassed by /b/ → [β] are necessarily a subset of the faithfulness violations amassed by /p/ → [β]. In particular, /b/ → [β] violates IDENT(cont) whereas /p/ → [β] violates IDENT(cont) as well as a second faithfulness constraint, IDENT(voice). Even in a framework with weighted constraints, it is logically the case that the penalty for violating IDENT(cont) + IDENT(voice), as in /p/ → [β], will always be equal or greater to the penalty for violating only IDENT(cont), as in /b/ → [β]. Therefore, it is not possible to have a case where /p/ → [β] is favored relative to /b/ → [β]. In section 2.4.3, this line of reasoning is laid out in more detail.

To allow saltation, it must be <u>possible</u> for a short journey (e.g., /b/ → [β]) to incur a greater penalty than a long journey (e.g., /p/ → [β]). As a solution, I adopt the *M_AP family of

---

[24] Software developed by Colin Wilson and Ben George, made available for public use by Bruce Hayes at http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

faithfulness constraints, proposed by Zuraw (2007). Unlike traditional IDENT constraints, *MAP

constraints are not restricted to cases involving the change of a single feature; instead, they

penalize correspondences between any two natural classes of sounds. The constraints are

formalized as follows, adopted from Zuraw (2007):

(33) ***MAP formalized***
           *MAP(x, y): violated when a sound that is a member of natural class x corresponds to a
             sound that is a member of natural class y.[25]

For the purposes of the cases considered here, what will be necessary is segment-specific

versions of the constraints. For instance, *MAP(p, β) would be violated whenever [p] is in

correspondence with [β].[26] In this case, the constraint *MAP(p, β) may be considered notational

shorthand for *MAP($\begin{bmatrix} -\text{voice} \\ -\text{cont} \\ +\text{labial} \end{bmatrix}$, $\begin{bmatrix} +\text{voice} \\ +\text{cont} \\ +\text{labial} \end{bmatrix}$), where each of the corresponding natural classes

happens to be made up of only a single segment.

    It is worth noting that these *MAP constraints are not inconsistent with traditional

faithfulness constraints. Indeed, traditional faithfulness constraints can be treated as special cases

of *MAP constraints—for instance, *MAP([−voice], [+voice]) would be violated whenever a

voiceless sound is in correspondence with a voiced sound. Likewise, *MAP(C, ∅) would be

---

[25] Zuraw's formalism also specifies particular contexts in which the pair of sounds must not be in correspondence
(e.g., a sound of natural class x in context A__B should not correspond to sound of natural class y in context C__D).
The context-specific version of the constraints is not necessary here, so I stick to this context-free version for
simplicity.

[26] Hypothetically, any correspondence relationship (i.e., input-output, output-output, base-reduplicant) is possible,
but an input-output correspondence is conceptually odd in this case because the constraints are intended to be
sensitive to the relative similarity of the sounds involved. It is unclear how to judge the similarity between an
abstract input form and a surface form. My analysis of saltation is fully consistent with an output-output
interpretation of the constraints, as is discussed further in section 4.6.3.

violated whenever a consonant is in correspondence with zero, making it equivalent to MAX-C. Thus, segment-specific faithfulness constraints and traditional faithfulness constraints can be unified into the same family of constraints.

Adopting the family of *MAP constraints, we see that even (otherwise) classical OT straightforwardly allows saltation. The solution for the Campidanian Sardinian case, where /p/ → [β] but /b/ remains [b], is shown in the tableaux in (34). The markedness constraints *V[−cont]V and *V[−voice]V are ranked above *MAP(p, β) so that underlying /p/ will change to [β]. *MAP(b, β) can then be ranked above *V[−cont]V so that /b/ is protected from changing.

(34) *Deriving saltation in OT with *MAP constraints*

  a) */p/ → [β]*

/VpV/	*MAP(b, β)	*V[−cont]V	*V[−voice]V	*MAP(p, β)	*MAP(p, b)
☞ VβV				*	
VbV		*!			*
VpV		*!	*		

  b) */b/ → [b]*

/VbV/	*MAP(b, β)	*V[−cont]V	*V[−voice]V	*MAP(p, β)	*MAP(p, b)
VβV	*!				
☞ VbV		*			

In sum, without expanding the constraint set beyond the types of faithfulness constraints allowed in classical OT, it is not possible to generate saltation. In the framework used here, *MAP constraints make it possible for large changes to preferred over small changes, which is

103

essential for deriving saltations.[27] However, allowing saltation is only one half of the puzzle.

Next, I will discuss how the theory can be constrained.

4.2.4 Constraining the theory: The P-map bias

*Map constraints are clearly much more powerful than traditional faithfulness constraints –

indeed, the set of *Map constraints can do what traditional faithfulness constraints can do and

much more. Thus we should be concerned about the implications of adding such a powerful tool

to our phonological framework. For instance, one may wonder, if any *Map constraint can be

freely ranked, then why is there not a preponderance of completely arbitrary patterns across the

world's languages? Why is it unlikely that we will find a yet undiscovered Language X in

which /β/ changes to [l], /r/ changes to [f], /q/ changes to [m], and /t/ changes to [ɢ], all before

vowels?[28]

The answer is that the constraints are not freely ranked. Zuraw (2007) argues that the set of

*Map constraints are initially ranked on the basis of Steriade's P-map, or perceptibility map

(Steriade, 2001/2008). The P-map represents knowledge that speakers have about the relative

perceptual distance between any two pairs of sounds in a given phonological context. In addition,

learners are claimed to have a minimal modification bias – phonological changes are assumed by

default to affect sounds in a way that results in the smallest perceptual change. Following

Steriade, Zuraw proposes that the knowledge encoded in the P-map is translated into *a priori*

---

[27] Constraint conjunction is another way of updating the set of possible constraints to allow saltation (e.g., Lubowicz, 2002; Ito & Mester, 2003). Problems with the constraint conjunction approach are discussed in section 2.5.1.

[28] There are, of course, cases of unnatural/arbitrary patterns that occur in languages (e.g., Hellberg, 1978; Anderson, 1981), and these patterns must be learnable at some level. Indeed, saltation is an example of pattern that, as I argue in this dissertation, is dispreferred but learnable. It is therefore desirable that our phonological theory have the ability to account for the arbitrary but learnable patterns when they do arise.

constraint rankings. Thus, *MAP constraints penalizing correspondences between perceptually

dissimilar sounds (in a given context) are ranked higher than constraints penalizing

correspondences between similar sounds. This *a priori* ranking represents the default rankings

for the *MAP constraints, but the default hierarchy can be overturned if contradicted by sufficient

evidence in the learner's language input.

I adopt Zuraw's proposal that the *MAP constraints are constrained by a substantive bias

based on the P-map. In the MaxEnt framework adopted here, I follow Wilson (2006) in

implementing the bias computationally via the prior (though not in the details of implementation;

see section 4.6.2). Instead of the constraints having an *a priori* default ranking, they will be

assigned individual *a priori* preferred weights (in the model, each constraint will have a different

μ). Intuitively, these weights bias the learner to believe that changes between similar sounds are

more likely than changes between dissimilar sounds, consistent with Steriade's principle of

minimal modification.

Based on the difference in relative similarity between the sounds involved, *MAP(p, β) will

have a higher preferred weight than *MAP(b, β). Of course, as we saw above in (34), to get a

saltation it must be possible to subvert this preferred hierarchy; that is, it must be possible for

MAP(b, β) to attain a higher weight than *MAP(p, β) despite the default P-map bias. A virtue of

the MaxEnt framework is that constraint weights can gradually shift away from their prior values

during the learning process as the model receives input data.

Thus, the general idea is that the prior, based on the P-map, makes saltations difficult to

learn, but with enough training, the prior weights can shift such that a grammar containing

saltation is ultimately possible.

4.3 Implementing the bias via the prior

I implemented three versions of the model: one with a substantive learning bias based on the P-map, one with a completely flat prior (all constraints have a default weight of 0), and one with no substantive bias, but a general preference for non-alternation. The latter two models will serve as comparisons for the substantively biased model. The basic architecture of the three models is the same. The only difference is how the prior is set in each model.

4.3.1 Substantively biased model

The biased version of the model was implemented by assigning each *MAP constraint an individual preferred weight (μ) based on the perceptual similarity of the pair of sounds specified in the constraint. Note that this is different than the way Wilson (2006) implemented the substantive bias in his model; a comparison of our two approaches is provided in section 4.6.2. For my implementation, I needed to consider two main issues: (1) how to define perceptual similarity, and (2) how to generate the μ for each constraint based on that similarity. I will discuss these in turn.

First, determining how to define and measure perceptual similarity is not trivial. In reality, listeners probably take many factors into account when making such judgments (e.g., see Steriade 2001/2008; Mielke, 2012; Cristia et al., in press). I use confusability as a measure of perceptual similarity, where the confusability of two speech sounds is determined according to the results of standard identification experiments (e.g., Miller & Nicely, 1955; Singh & Black, 1966; Wang & Bilger, 1973; Cutler et al., 2004). In these experiments, participants listen to recordings of speech sounds (with or without noise) and identify which sound they heard in some target location. A confusion matrix can then be calculated based on the responses recorded for

106

each type of sound. Even if somewhat coarse, confusability is a straightforward way of approximating perceptual similarity that works well for the purposes of this study.

In particular, I summed the confusion values from Tables 2 and 3 of Wang and Bilger 1973, where the target consonants were placed in CV (Table 2) and VC (Table 3) contexts. The stimuli were also presented in noise, and the values in these two tables were the summed values across all signal-to-noise ratios. Reasons for using these values, and potential implications of doing so, are discussed further in section 4.4.5 below.

The second consideration is how to go from the confusion probabilities (from perception experiments) to the preferred weights for the prior. To do so, the confusion values were entered into a separate MaxEnt model intended only to generate the prior weights. Intuitively, one can think of this model as representing the learner's experience perceiving speech sounds. For reference, the confusion values entered in the model are provided in Table 5. Each relevant *MAP constraint was also included in the model, and violations were marked whenever the two sounds listed in the constraint were confused for one another. For instance, a violation was marked for *MAP(p, v) when [p] was confused for [v], or vice versa.[29]

Table 5. Confusion values for the combined CV and VC contexts from Wang & Bilger (1973, Tables 2 and 3), which were used to generate the prior. Only the sound pairs relevant for the current study are shown here.

Stimulus	Responses				Stimulus	Responses			
	p	b	f	v		t	d	θ	ð
p	1844	54	159	26	t	1765	107	92	26
b	206	1331	241	408	d	91	1640	75	193
f	601	161	1202	93	θ	267	118	712	135
v	51	386	127	1428	ð	44	371	125	680

---

[29] For the prior, μ was set to 0 and σ^2 was set to 10,000. This value of σ^2 is sufficiently high that the weights were essentially free to be whatever they needed to be in order to best match the confusion probabilities in the input data.

The result is that the *MAP constraints received weights based on how often the two sounds named in the constraint were confused with each other in the confusion experiment. Recall that mutual confusability is taken here as a rough approximation of perceptual similarity. The resulting weights are provided below in Table 6. Sounds that are very confusable with each other, and thus assumed to be highly similar, received low weights whereas sounds that are dissimilar received more substantial weights. For instance, [b] and [v] are very similar to each other, so *MAP(b, v) received the small weight of 1.30. On the other hand, [p] and [v] are quite dissimilar, so *MAP(p, v) received a weight of 3.65.[30] These weights were entered directly into the primary learning model as the preferred weights (μ) for each constraint in the prior. It is ideal to derive the prior weights directly from the confusion data in a systematic way, as I have done here, without direct manipulation of the weights by the modeler. Doing so arguably provides a better test of the model, which allows us to draw better conclusions about how the success (or failure) of the model relates to the relationship between perceptual similarity (defined in terms of confusability) and the learning process.

For the version of the model used in the next section, σ^2 was set to 0.6 for every constraint. Other values of σ^2 are considered in section 4.4.4 below.

Table 6. Prior weights (μ) for *MAP constraints based on confusion data.

Labial sounds		Coronal sounds	
Constraint	Prior weight (μ)	Constraint	Prior weight (μ)
*MAP(p, v)	3.65	*MAP(t, ð)	3.56
*MAP(f, v)	2.56	*MAP(θ, ð)	1.91
*MAP(p, b)	2.44	*MAP(t, d)	2.73
*MAP(f, b)	1.96	*MAP(θ, d)	2.49
*MAP(p, f)	1.34	*MAP(t, θ)	1.94
*MAP(b, v)	1.30	*MAP(d, ð)	1.40

---

[30] The fact that sounds were sometimes misheard means that the weights were forced to stay near 0 (relatively speaking), even though σ^2 was set high. If the weights strayed too far from 0, then confusions would be predicted by the model too rarely.

4.3.2 Unbiased model

For comparison, the second version of the model, which I will call the "unbiased" model, did <u>not</u> have a substantive bias. The unbiased model had a "flat" prior: every constraint had the same μ (set to 0, where 0 means the constraint has no effect on the outcome) and σ^2 (set to 0.6, i.e. the same value as in the biased model).[31] It was otherwise identical to the biased model.

4.3.3 High faith model

In the previous unbiased model, the μ for every constraint was set to 0, but in the substantively bias model, each *MAP constraint has a non-zero weight. Thus, the unbiased model may not be the fairest comparison for the biased model because it differs from the biased model on two accounts: not having a substantive bias <u>and</u> having faithfulness default to 0. When we begin evaluating the model predictions below, we will see that this turns out to be important.

To deal with this problem, I created a third model, which I will call the "high faith model," in which every *MAP constraint is assigned a prior weight of 2.27. This value is the mean of all the *MAP weights that make up the prior of the substantively biased model. The mean of the prior weights in the biased model was chosen in order to give the model the best chance of succeeding. In all other ways, the model is identical to the other two models. The high faith model is similar to the biased model in that all of the *MAP constraints have non-zero weights, but unlike the substantively biased model, those weights do not vary according to perceptual similarity.

---

[31] In principle, a better comparison might involve fitting the unbiased model (as well as the high faith model) to the σ^2 that maximizes its own performance rather using the same σ^2 that maximized the performance of the biased model. In practice, this turns out not to matter much; the unbiased model and the high faith model never reach the level of performance as the biased model, regardless of the σ^2 used (see Figure 11). Other values of σ^2 are considered in section 4.4.4.

The *MAP constraints can be thought of as output-output faithfulness constraints (Benua, 1997) or paradigm uniformity constraints (Hayes, 1997; Steriade, 2000). Thus, having a non-zero prior for these constraints can be considered as a default preference to avoid alternation. This idea is discussed in more detail in section 4.6.3 below.

4.4 Testing the model

To test the MaxEnt model's performance, the model was provided the same training data received by the experimental participants (see sections 3.4 and 3.5). The model predictions were then fitted to the aggregate experimental results.

During learning, the model considered all obstruents within the same place of articulation as possible outputs for a given input. For instance, for input /p/, the model considered the set of {[p], [b], [f], [v]} as possible outputs. In fact, each input had only one possible output in the actual observed training data because there was no free variation in the experimental input (e.g., in Experiment 1, /p/ or /b/ (depending on condition) changed to [v] 100% of the time). Thus, during learning the model was trying to account for the fact that the winning output was the winner and the other three possible outputs were losers. At test, the model only considered the relative probability of two possible outputs – that is, the two outputs that the experimental participants considered. The goal was to put the model and the participants in the same situation: during training, neither knew what the test was going to be like, so they had to consider all possibilities. But at test, both were forced to choose between only two possible outcomes.

I focus on the cases in Experiment 1 and in the Stops sub-group of Experiment 2.[32] Table 7

provides an overview of the training data for these cases. Overall, the comparison with

Experiment 1 illustrates that the model predicts a large amount of generalization to intermediate

sounds (as in the Potentially Saltatory condition), but less generalization to nearby sounds when

those sounds are <u>not</u> intermediate (as in the Control condition). Further, the comparison with

Experiment 2 shows that the model predicts difficulty learning that intermediate sounds do not

change when trained on explicit saltations, consistent with the experimental results. In the

following sections, I consider the results in detail.

Table 7. Overview of training data for the MaxEnt model.

Experiment 1		Experiment 2 – Stops sub-group	
Potentially Saltatory condition	Control condition	Saltatory condition	Control condition
18 p → v	18 b → v	18 p → v	18 b → v
18 t → ð	18 d → ð	18 t → ð	18 d → ð
		9 b → b	9 p → p
		9 d → d	9 t → t

4.4.1 Experiment 1: Dispreference for saltation with ambiguous input

Recall that in Experiment 1, participants were trained on potentially saltatory alternations

(e.g., [p ~ v]) and then tested on intermediate sounds that they had never seen before (e.g., [b]

and [f]). By also changing intermediate sounds, participants could avoid the saltatory system (in

favor of a neutralizing system: e.g., both [p] and [b] changed to [v]), but doing so would require

them to posit new alternations without evidence. On the other hand, they could avoid positing

new alternations, but doing so would render the phonological system saltatory. At test,

---

[32] The Fricatives sub-group of Experiment 2 was not considered here because there was an additional complication in the results likely due to the fact that [θ] and [ð] have the same orthographic representation in English (see section 3.5.3). The model implemented here is not equipped to deal with orthographic effects.

participants changed the intermediate sounds at a very high rate, choosing the anti-conservative option, thereby avoiding the saltation. When they were instead trained on alternations between similar sounds (e.g., [b ~ v]) and tested on novel sounds that were not intermediate (e.g, [p] and [f]), participants changed the novel sounds much less frequently at test, indicating that saltation avoidance, and not just generalization to nearby sounds, was playing a role in the effect.

4.4.1.1 Substantively biased model

Table 8 shows how the constraint weights changed from their prior weights as a result of the training data in Experiment 1. In the Potentially Saltatory condition (training = p → v; t → ð), we see that both markedness constraints, *V[–voice]V and *V[–cont]V, pick up weights in order to motivate voiceless stops changing into voiced fricatives. Likewise, the weights for *MAP(p, v) and *MAP(t, ð) are substantially reduced because the training data provide evidence that it is indeed acceptable for those sounds to be mapped to one another (contrary to the P-map). Other *MAP constraints involving [p] and [t] (i.e., *MAP(p, b), *MAP(t, d), *MAP(p, f), *MAP(t, θ)) have modest increases in their weights because they all play a role (during learning though not at test, where only two outcomes are possible) in ensuring that /p/ and /t/ are mapped to [v] and [ð], respectively, rather than to some other sound (i.e., [b], [f], [d], or [θ]). Weights for the remaining *MAP constraints remain at their prior weights because they have no effect on the /p/ → [v] or /t/ → [ð] outcomes.

In the Control condition (training = b → v; d → ð), the markedness constraint *V[–cont]V gets a substantial increase in weight to motivate spirantization. The markedness constraint *V[–voice]V receives a small weight due only to its limited role in keeping /b/ and /d/ from changing into [p] and [t], respectively. *MAP(b, v) and *MAP(d, ð) have substantially reduced

112

weights (almost to 0) because the training data indicates that such mappings are acceptable. Once again, other *MAP constraints involving [b] or [d] receive modest increases in their weights so that /b/ and /d/ will be mapped to [v] and [ð] rather than some other sound. The weights of the remaining *MAP constraints, which have no effect on the outcomes of the training data, do not change from their prior values.

Table 8. Prior constraint weights and post-learning weights in the Potentially Saltatory and Control conditions of Experiment 1 (substantively biased model).

Constraint	Prior weight	Post-learning weight	
		Potentially Saltatory condition	Control condition
*V[–voice]V	0	2.20	0.57
*V[–cont]V	0	1.86	1.80
*MAP(p, v)	3.65	2.17	3.65
*MAP(t, ð)	3.56	2.22	3.56
*MAP(p, b)	2.44	2.77	2.48
*MAP(t, d)	2.73	3.02	2.76
*MAP(p, f)	1.34	1.90	1.34
*MAP(t, θ)	1.94	2.34	1.94
*MAP(b, v)	1.30	1.30	0.15
*MAP(d, ð)	1.40	1.40	0.25
*MAP(f, v)	2.56	2.56	2.56
*MAP(θ, ð)	1.91	1.91	1.91
*MAP(b, f)	1.96	1.96	2.25
*MAP(d, θ)	2.49	2.49	2.70

From these weights, the model calculates the predicted probability of each output candidate at test given each input. These probabilities, which are calculated as described in section 4.2.1, can be represented in OT tableaux. A couple of examples are given in (35) for the Potentially Saltatory condition of Experiment 1. Constraint weights are taken from Table 8.

(35) *Calculating predicted probabilities in tableaux*

a) *Input /p/ in Experiment 1, Potentially Saltatory condition*

/VpV/	*V[−voice]V 2.20	*MAP(p, v) 2.17	*V[−cont]V 1.86	*MAP(b, v) 1.30	Penalty score	$e^{(-\text{penalty})}$	Predicted outcome
VvV		2.17			2.17	.1142	.87
VpV	2.20		1.86		4.06	.0172	.13

b) *Input /b/ in Experiment 1, Potentially Saltatory condition*

/VbV/	*V[−voice]V 2.20	*MAP(p, v) 2.17	*V[−cont]V 1.86	*MAP(b, v) 1.30	Penalty score	$e^{(-\text{penalty})}$	Predicted outcome
VvV				1.30	1.30	.2725	.64
VbV			1.86		1.86	.1557	.36

The predicted probabilities of each output candidate can then be compared to the experimental results. Table 9 shows the model predictions along with the aggregate experimental results for Experiment 1.

The model correctly predicted that the alternations presented during training would be successfully learned, as reflected by the high percentages in the shaded rows. Of greater interest, however, is that the model correctly predicted the two significant effects related to the untrained sounds. First, the model predicted that untrained sounds would be changed more often when they were intermediate (i.e., Potentially Saltatory condition) than when they were not intermediate (Control condition). In other words, the model predicted saltation avoidance when presented with ambiguous data like in Experiment 1.

Second, the model predicted a preference for changing intermediate stops (i.e., [b ~ v] and [d ~ ð]) compared to changing intermediate fricatives ([f ~ v] and [θ ~ ð]). This difference falls out directly from the prior based on the P-map. Taking the labials as an example: because [b] and

114

[v] are more perceptually similar than [f] and [v], *MAP(b, v) has a lower prior weight than *MAP(f, v). As a result, changing [f] to [v] garners a greater penalty than changing [b] to [v], leading to the preference observed. Recall that abstract features are not able to account for this difference because both [b] and [f] differ from [v] by a single feature.

Table 9. Model predictions (biased model) and experimental results from the Potentially Saltatory and Control conditions from Experiment 1. Values represent percentage of trials in which the changing option was chosen (experiment) or in which the changing option was predicted (model). Shaded rows indicate the trained alternations.

Experiment 1: Potentially Saltatory condition					
Labials			Coronals		
	Model prediction	Experimental result		Model prediction	Experimental result
p → v	87	98	t → ð	86	95
b → v	64	73	d → ð	61	67
f → v	41	49	θ → ð	57	41

Experiment 1: Control condition					
Labials			Coronals		
	Model prediction	Experimental results		Model prediction	Experimental results
b → v	84	88	d → ð	82	89
p → v	22	18	t → ð	23	23
f → v	12	14	θ → ð	21	18

4.4.1.2 Unbiased model

For comparison, Table 10 shows how the weights for the unbiased model change in these two conditions. Recall that the constraints all start with a prior weight of 0 in the unbiased instantiation of the model. As Table 10 clearly shows, the faithfulness constraints have little reason to change based on training in this model. A few of the *MAP constraints pick up a modest weight; these constraints play a small role in avoiding alternations not seen during

training (e.g., ensuring that [p] → [v], not [p] → [f] or [b]). Most of the work in the model is done by the markedness constraints alone.

Table 10. Prior constraint weights and post-learning weights in the Potentially Saltatory and Control conditions of Experiment 1 (unbiased model).

| | | Post-learning weight | |
| | | Potentially Saltatory | |
Constraint	Prior weight	condition	Control condition
*V[–voice]V	0	1.49	1.41
*V[–cont]V	0	1.49	1.72
*MAP(p, v)	0	0	0
*MAP(t, ð)	0	0	0
*MAP(p, b)	0	0.54	0.15
*MAP(t, d)	0	0.54	0.15
*MAP(p, f)	0	0.54	0
*MAP(t, θ)	0	0.54	0
*MAP(b, v)	0	0	0
*MAP(d, ð)	0	0	0
*MAP(f, v)	0	0	0
*MAP(θ, ð)	0	0	0
*MAP(b, f)	0	0	0.56
*MAP(d, θ)	0	0	0.56

The unbiased model's predictions resulting from this grammar are given in Table 11. The most noticeable aspects of the model's predictions is that they are identical for labials and coronals, and that they are identical for voiced stops and voiceless fricatives in the Potentially Saltatory condition. Because the model does not have access to perceptual similarity, it does not differentiate between pairs of sounds that differ in the same features. Crucially, the model is unable to account for the significant difference in how often voiced stops and voiceless fricatives were changed in the Potentially Saltatory condition. From the unbiased model's perspective, there is no *a priori* difference between the alternations [b ~ v] and [f ~ v], for instance; they both differ by a single feature.

116

The second striking aspect of the model's predictions is how much it overestimates the amount of generalization to untrained sounds in the Control condition. This extreme amount of overgeneralization is due to the fact that the *MAP constraints have a prior weight of 0. Because they are absent in training, there is no reason to raise their weights, so they remain at 0. The extreme nature of this overgeneralization is not so much a problem of lacking a substantive bias, but rather a problem of having *MAP constraints set with a prior weight of 0. Therefore, this unbiased model may not be the best test of how much work the substantive bias is doing in the model. With this in mind, I now turn to an instantiation of the unbiased model with high faith.

Table 11. Model predictions (unbiased model) and experimental results from the Potentially Saltatory and Control conditions from Experiment 1. Values represent percentage of trials in which the changing option was chosen (experiment) or in which the changing option was predicted (model). Shaded rows indicate the trained alternations.

Experiment 1: Potentially Saltatory condition					
Labials			Coronals		
	Model prediction	Experimental result		Model prediction	Experimental result
p → v	95	98	t → ð	95	95
b → v	82	73	d → ð	82	67
f → v	82	49	θ → ð	82	41

Experiment 1: Control condition					
Labials			Coronals		
	Model prediction	Experimental results		Model prediction	Experimental results
b → v	85	88	d → ð	85	89
p → v	96	18	t → ð	96	23
f → v	80	14	θ → ð	80	18

4.4.1.3 High faith model

Table 12 illustrates how the weights for the high faith model change when presented with the training data from these two conditions. The behavior of the weights is similar to the weights

in the substantively biased model. The markedness constraint have weights that increase while the *MAP penalizing the alternating sounds in training have weights that decrease. The other *MAP constraints have either small modifications to their weights or no change in their weights (if they do not affect the outcome at all).

Table 12. Prior constraint weights and post-learning weights in the Potentially Saltatory and Control conditions of Experiment 1 (high faith model).

Constraint	Prior weight	Post-learning weight	
		Potentially Saltatory condition	Control condition
*V[–voice]V	0	1.62	2.19
*V[–cont]V	0	1.62	0.75
*MAP(p, v)	2.27	1.22	2.27
*MAP(t, ð)	2.27	1.22	2.27
*MAP(p, b)	2.27	2.51	2.32
*MAP(t, d)	2.27	2.51	2.32
*MAP(p, f)	2.27	2.51	2.27
*MAP(t, θ)	2.27	2.51	2.27
*MAP(b, v)	2.27	2.27	0.85
*MAP(d, ð)	2.27	2.27	0.85
*MAP(f, v)	2.27	2.27	2.27
*MAP(θ, ð)	2.27	2.27	2.27
*MAP(b, f)	2.27	2.27	2.60
*MAP(d, θ)	2.27	2.27	2.60

The high faith model's predictions for Experiment 1 are given in Table 13. Once again, the model's inability to differentiate the labials and coronals is exhibited by its predictions. Moreover, in the Potentially Saltatory condition, it still cannot account for the observed preference for changing intermediate voiced stops over changing intermediate voiceless fricatives. These problems are both due to the fact that all of the *MAP constraints have the same prior weight.

The critical problem with this model is that it predicts the wrong direction for when learners will generalize to untrained stops. In the Potentially Saltatory condition, it predicts too little

118

generalization to intermediate voiced stops (only 34% are predicted to change, compared with 67–73% in the experiment). By contrast, it predicts too much generalization to untrained voiceless stops in the Control condition (66% predicted to change, compared with 18–23% in the experiment). In other words, the model predicts that learners will generalize more often to untrained stops that are not intermediate (Control condition) than to untrained stops that are intermediate (Potentially Saltatory condition), which is the opposite of the actual experimental results. This problem is also due to the fact that the *MAP constraints have the same prior weights. The model has no way of knowing *a priori* that [b] → [v] is a more plausible alternation than [p] → [v], which is necessary to get the correct pattern of generalization.

In sum, we can conclude that the substantive bias is crucial in order for the model to account for two aspects of the experimental results: the basic saltation effect (i.e., more generalization to intermediate sounds than non-intermediate sounds) and the preference for changing the intermediate voiced stops more than the intermediate voiceless fricatives.

Table 13. Model predictions (high faith model) and experimental results from the Potentially Saltatory and Control conditions from Experiment 1. Values represent percentage of trials in which the changing option was chosen (experiment) or in which the changing option was predicted (model). Shaded rows indicate the trained alternations.

Experiment 1: Potentially Saltatory condition					
Labials			Coronals		
	Model prediction	Experimental result		Model prediction	Experimental result
p → v	88	98	t → ð	88	95
b → v	34	73	d → ð	34	67
f → v	34	49	θ → ð	34	41

Experiment 1: Control condition					
Labials			Coronals		
	Model prediction	Experimental results		Model prediction	Experimental results
b → v	79	88	d → ð	79	89
p → v	66	18	t → ð	66	23
f → v	18	14	θ → ð	18	18

4.4.2 Experiment 2: Difficulty learning saltations with explicit evidence

When participants were explicitly presented with saltation during training (Experiment 2), recall that they had difficulty learning that intermediate sounds did not change. For instance, when learning the saltatory alternation [p ~ v] with unchanging [b], they had a tendency to incorrectly change intermediate [b] at test (Saltatory condition). But when learning comparable alternations that were not saltatory, such as [b ~ v] with unchanging [p], they rarely changed [p] in error (Control condition).

4.4.2.1 Substantively biased model

Table 14 shows the model predictions along with the actual experimental results for the training data in Experiment 2 (Stops sub-group). The model correctly predicts a tendency to make errors on intermediate [b] in the Saltatory condition. By contrast, in the Control condition, the model predicts very few errors, consistent with the actual results.

In addition, the model correctly predicts that that <u>untrained</u> fricatives ([f] and [θ]) will be changed more often when they are intermediate (Saltatory condition) than when they are not intermediate (Control condition), replicating the basic effect from Experiment 1.

Table 14. Model predictions (biased model) and experimental results from the Saltatory and Control conditions from Experiment 2, Stops sub-group (in percentage changing option chosen/predicted). Shaded rows represent trained cases. Asterisks mark the places where participants (and the model) were trained on the opposite of what is shown (e.g., b → v * means that participants were trained on b → b, but the value shown represents the percentage of times that participants (or the model) changed b → v in spite of the training).

		Experiment 2: Saltatory condition					
		Labials				Coronals	
		Model prediction	Experimental result			Model prediction	Experimental result
p → v		82	95	t → ð		82	93
b → v	*	27	21	d → ð	*	26	21
f → v		47	52	θ → ð		63	54

		Experiment 2: Control condition					
		Labials				Coronals	
		Model prediction	Experimental results			Model prediction	Experimental results
b → v		75	91	d → ð		75	97
p → v	*	7	9	t → ð	*	8	4
f → v		8	14	θ → ð		14	29

Table 15 demonstrates how the prior constraint weights change in response to the training data of Experiment 2 (Stops sub-group), resulting in the predictions given above. Overall, the changes in weights are similar to those seen for Experiment 1 (Table 8). In the Saltatory condition (training = p → v; b → b; t → ð; d → d), the weights of both markedness constraints, *V[–voice]V and *V[–cont]V, are increased to motivate the alternations between voiceless stops and voiced fricatives. The weights of *MAP(p, v) and *MAP(t, ð) are substantially reduced due to the evidence in training that those sounds alternate. Crucially, the weights of *MAP(b, v) and *MAP(d, ð) are bolstered to protect the intermediate stops from changing; however, because the prior weights of these constraints were low due to the similarity of the sounds involved, their weights are not bolstered enough to fully protect the intermediate sounds. The model therefore predicts errors on intermediate sounds at a rate that is comparable to what was found in

Experiment 2. Other constraints involving [p], [b], [t], or [d] each receive a modest increase in their weights for their role in avoiding other alternations not seen during training.

In the Control condition (training = b → v; p → p; d → ð; t → t), only the markedness constraint *V[–cont]V receives a substantial boost to its weight in order to motivate spirantization of /b/ and /d/ whereas *MAP(b, v) and *MAP(d, ð) both see their weights reduced to 0 in order to permit the alternations seen during training. The weights of *MAP(p, v) and *MAP(t, ð) are increased to protect /p/ and /t/ from changing, but the increase is very small because the prior weights of those constraints are already quite high. In this case, the training data and the prior support the same conclusion: no [p ~ v] alternations and no [t ~ ð] alternations. Thus, very few errors are predicted on these sounds, consistent with the experimental results.

Table 15. Prior constraint weights and post-learning weights (biased model) in the Saltatory and Control conditions of Experiment 2 (Stops sub-group).

Constraint	Prior weight	Post-learning weight	
		Saltatory condition	Control condition
*V[–voice]V	0	2.45	0.13
*V[–cont]V	0	1.05	1.12
*MAP(p, v)	3.65	1.96	3.79
*MAP(t, ð)	3.56	2.01	3.72
*MAP(p, b)	2.44	2.94	2.65
*MAP(t, d)	2.73	3.16	2.91
*MAP(p, f)	1.34	1.74	2.03
*MAP(t, θ)	1.94	2.21	2.45
*MAP(b, v)	1.30	2.02	0
*MAP(d, ð)	1.40	2.09	0
*MAP(f, v)	2.56	2.56	2.56
*MAP(θ, ð)	1.91	1.91	1.91
*MAP(b, f)	1.96	2.02	2.29
*MAP(d, θ)	2.49	2.53	2.71

4.4.2.2 Unbiased model

Table 16 shows the unbiased model's predictions for Experiment 2. Like in Experiment 1, we see that the model is unable to account for the basic saltation effect. Specifically, it predicts many more errors on voiceless stops in the Control condition (52% errors predicted vs. 4–9% errors in the actual experiment) than on intermediate voiced stops in the Saltation condition (37% errors predicted vs. 21% errors in the actual experiment). The model also overgeneralizes for untrained fricatives, just as it did for Experiment 1.

Table 16. Model predictions (unbiased model) and experimental results from the Saltatory and Control conditions from Experiment 2, Stops sub-group (in percentage changing option chosen/predicted). Shaded rows represent trained cases. Asterisks mark the places where participants (and the model) were trained on the opposite of what is shown (e.g., b → v * means that participants were trained on b → b, but the value shown represents the percentage of times that participants (or the model) changed b → v in spite of the training).

Experiment 2: Saltatory condition						
Labials				Coronals		
	Model prediction	Experimental result			Model prediction	Experimental result
p → v	91	95	t → ð		91	93
b → v *	37	21	d → ð *		37	21
f → v	88	52	θ → ð		88	54

Experiment 2: Control condition						
Labials				Coronals		
	Model prediction	Experimental results			Model prediction	Experimental results
b → v	66	91	d → ð		66	97
p → v *	52	9	t → ð *		52	4
f → v	56	14	θ → ð		56	29

A look at how the weights change in response to the training data (**Table 17**) reveals why the model fails. The reason for the overgeneralization on untrained sounds is the same as in Experiment 1: the relevant *MAP constraints (i.e., *MAP(f, v) and *MAP(θ, ð)) have a prior

weight of 0 and because they do not appear in the training data, there is never a reason for the weights to go up.

The reason that the model predicts too many errors for voiceless stops in the Control condition is subtler. Using the labials as an example, the only constraint that can motivate the [b] → [v] change found in the training data is *V[–cont]V; however, raising this constraint also motivates [p] to change. The only way to protect [p] from changing is to increase the weight of *MAP(p, v). But because all of the constraints start at 0, the model is unable to raise the weight of *V[-cont]V enough to motivate the [b] → [v] alternation while simultaneously raising the weight of *MAP(p, v) high enough to sufficiently protect [p] from changing. *MAP(p, v) would need to be raised quite a bit above *V[–cont]V in order to protect [p]. By contrast, in the biased model, *MAP(p, v) starts out with a high weight, so such an arrangement is possible.

Table 17. Prior constraint weights and post-learning weights (unbiased model) in the Saltatory and Control conditions of Experiment 2 (Stops sub-group).

		Post-learning weight	
Constraint	Prior weight	Saltatory condition	Control condition
*V[–voice]V	0	2.01	0.24
*V[–cont]V	0	0.36	0.65
*MAP(p, v)	0	0	0.82
*MAP(t, ð)	0	0	0.82
*MAP(p, b)	0	1.02	0.87
*MAP(t, d)	0	1.02	0.87
*MAP(p, f)	0	0.35	0.71
*MAP(t, θ)	0	0.35	0.71
*MAP(b, v)	0	0.89	0
*MAP(d, ð)	0	0.89	0
*MAP(f, v)	0	0	0
*MAP(θ, ð)	0	0	0
*MAP(b, f)	0	0.23	0.88
*MAP(d, θ)	0	0.23	0.88

4.4.2.3 High faith model

The predictions of the high faith model for Experiment 2 are given in Table 18. This model

fits the results of the Saltatory condition very well. However, like the regular unbiased model, it

also fails to predict the basic anti-saltation relationship between the two conditions. Specifically,

the model predicts more errors for the voiceless stops in the Control condition than for the

intermediate voiced stops in the Saltatory condition. The problem is not as pronounced as in the

unbiased model (20% errors for the high faith model vs. 52% errors for the unbiased model), but

the model *should* be predicting very few errors on the voiceless stops in the Control condition,

and the number should crucially be lower than the number of errors predicted for the voiced

stops in the Saltatory condition.

Table 18. Model predictions (high faith model) and experimental results from the Saltatory and Control conditions from Experiment 2, Stops sub-group (in percentage changing option chosen/predicted). Shaded rows represent trained cases. Asterisks mark the places where participants (and the model) were trained on the opposite of what is shown (e.g., b → v * means that participants were trained on b → b, but the value shown represents the percentage of times that participants (or the model) changed b → v in spite of the training).

Experiment 2: Saltatory condition						
Labials				Coronals		
	Model prediction	Experimental result			Model prediction	Experimental result
p → v	86	95	t → ð		86	93
b → v *	16	21	d → ð *		16	21
f → v	41	52	θ → ð		41	54

Experiment 2: Control condition						
Labials				Coronals		
	Model prediction	Experimental results			Model prediction	Experimental results
b → v	72	91	d → ð		72	97
p → v *	20	9	t → ð *		20	4
f → v	9	14	θ → ð		9	29

Looking at the behavior of the constraint weights (Table 19), we see that the reason for the failure once again stems from the model's lack of sensitivity to perceptual similarity. Because the prior weights for *MAP(b, v) and *MAP(p, v) are identical, the model has no way to motivate [b] → [v] without also motivating [p] → [v].

Table 19. Prior constraint weights and post-learning weights (high faith model) in the Saltatory and Control conditions of Experiment 2 (Stops sub-group).

Constraint	Prior weight	Post-learning weight	
		Saltatory condition	Control condition
*V[–voice]V	0	1.90	0
*V[–cont]V	0	1.02	1.31
*MAP(p, v)	2.27	1.10	2.70
*MAP(t, ð)	2.27	1.10	2.70
*MAP(p, b)	2.27	2.62	2.52
*MAP(t, d)	2.27	2.62	2.52
*MAP(p, f)	2.27	2.43	2.70
*MAP(t, θ)	2.27	2.43	2.70
*MAP(b, v)	2.27	2.68	0.39
*MAP(d, ð)	2.27	2.68	0.39
*MAP(f, v)	2.27	2.27	2.27
*MAP(θ, ð)	2.27	2.27	2.27
*MAP(b, f)	2.27	2.36	2.64
*MAP(d, θ)	2.27	2.36	2.64

4.4.3 Overall model performance

Taking all of the observations included in sections 4.4.1 and 4.4.2, we see that the predictions of the MaxEnt model with a substantive bias based on the P-map (implemented by way of the prior) produces an excellent overall fit to the experimental data ($r^2 = .95$). This is shown in Figure 8. Each point in the figure represents one observation (e.g., p → v in the Control condition of Experiment 1 would be one point), with the model prediction (x-axis) plotted against the aggregate experimental results across all participants (y-axis).

126

Figure 8. Predictions of the substantively biased model plotted against the experimental results, for each of the observations above. Overall $r^2 = .95$.
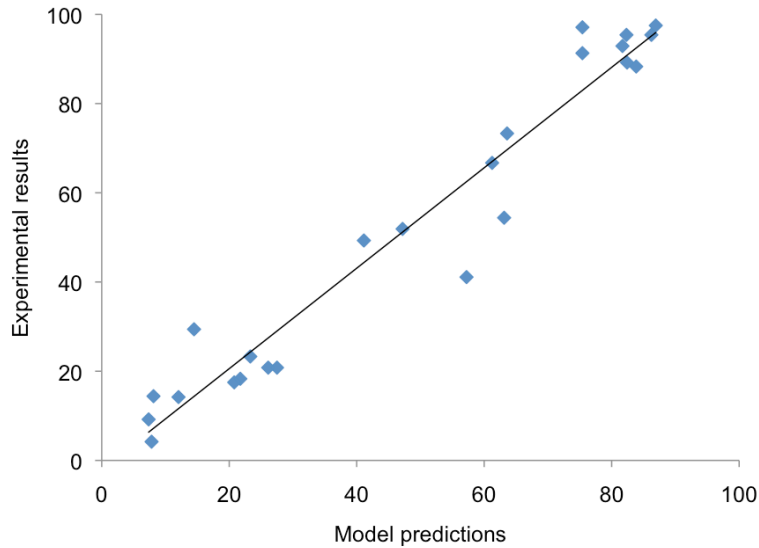
By comparison, the unbiased model results in a much poorer fit to the data ($r^2 =.25$), shown in

Figure 9. The high faith model (Figure 10) has a better fit to the data ($r^2 =.67$), but due to the

problems discussed above, its fit is still considerably worse than the fit of the substantively

biased model. In sum, having non-zero prior weights for the *MAP constraints results in a large

improvement in performance compared with a model in which all of the weights have 0 as a

default, but it is clear that having a substantive bias based on perceptual similarity results in

considerably better performance above and beyond just having non-zero weights for faithfulness.

Figure 9. Model predictions plotted against the experimental results, for the unbiased model. Overall $r^2 = .25$.
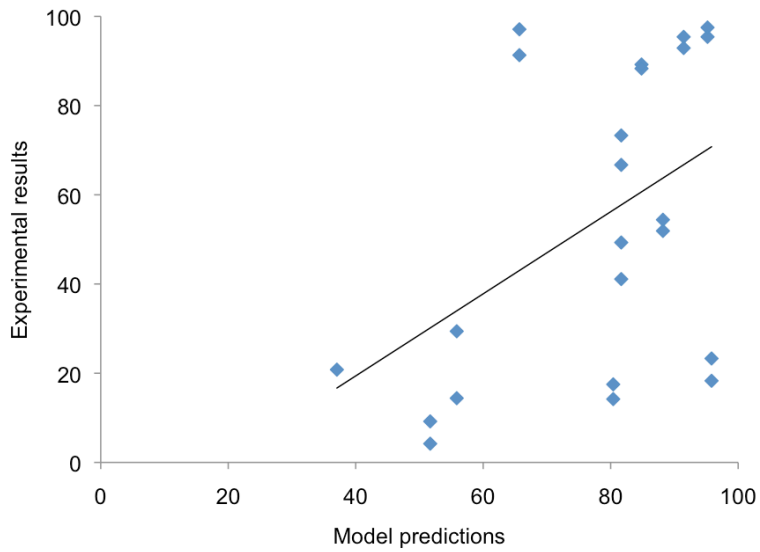
Figure 10. Model predictions plotted against the experimental results, for the high faith model. Overall $r^2 = .67$.
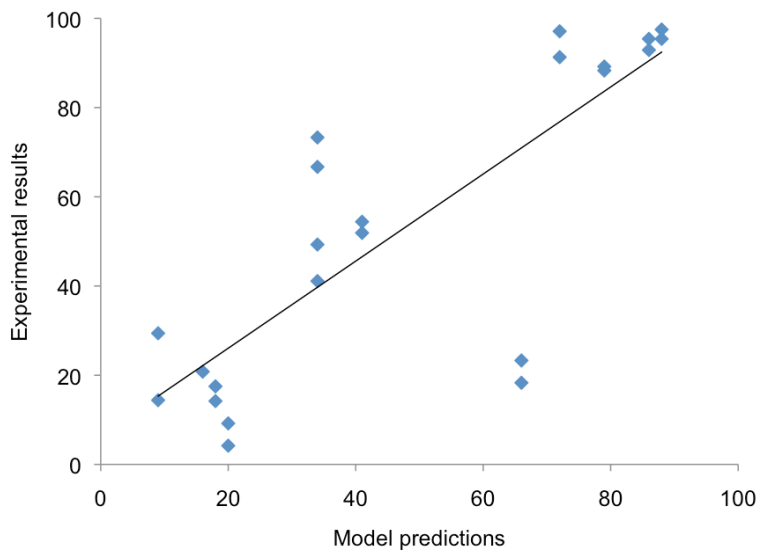
4.4.4 Effect of different σ^2 values

As implemented, the only free parameter in the model is the squared standard deviation, σ^2, of the prior distribution for each constraint. Recall the value of σ^2 determines how tightly

128

constraint weights are bound to their preferred weight (i.e., the μ of the prior distribution). Lower values of σ^2 means that more data are required to pull the weights away from μ, whereas higher values of σ^2 mean that the weights are freer to change in light of the training data.

I had no *a priori* assumptions about how to set σ^2, so several values for σ^2 were tested. To get the outputs reported in sections 4.4.1–4.4.3, σ^2 was set to 0.6, the value that maximized the proportion of variance explained by the model (r^2) when fitted to the experimental results. But it is worth considering how different values of σ^2 affect the model's performance.

Figure 11 shows the proportion of variance explained by the model (r^2) as a function of different values for σ^2, for all three models. The most striking aspect of the figure is that for all but the most extreme values of σ^2, the substantively biased model outperforms the high faith model by a considerable margin, and it outperforms the unbiased model by an even greater margin. Thus, the overall conclusion that the biased model outperforms the unbiased model and the high faith model is not dependent on choosing a precise value for σ^2.

Looking at the substantively biased model, we see that the model performs best between the σ^2 values of 0.5 and 0.7, the range at which r^2 reaches a virtual plateau around .94–.95. The reason is that these values of σ^2 represent the "Goldilocks" range that is "just right" (at least for this model): the values are low enough that the prior can still have a substantial effect on the outcome but high enough that the training data also have a substantial effect. As the value of σ^2 decreases from 0.5, we see that the model's performance begins to drop, with the drop becoming more abrupt as the value of σ^2 decreases to 0.2 and below. This decrease in performance occurs because as σ^2 drops, the prior becomes too strong such that the training data have little effect on the constraint weights. On the other side, as the value of σ^2 increases from 0.8 and beyond, the model's performance continues to decline at a gradual rate. At an extreme value of 100,000, the

prior has almost no practical effect on the constraint weights, leaving the weights to be almost

entirely dictated by the training data. As a result, all three models converge at (virtually) the

same predictions and thus have very similar r^2 values.

Figure 11. Proportion variance explained (r^2) by the substantively biased model, the high faith model, and
the unbiased model, according to the value of σ^2.
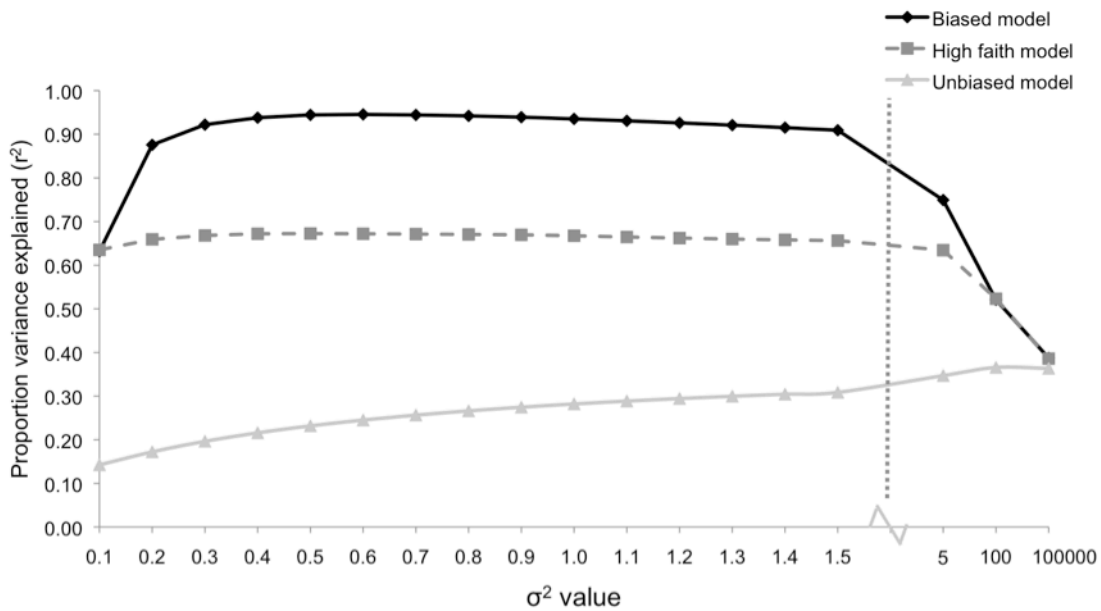
4.4.5 Effect of different confusion matrices used to derive the prior

Given that the prior weights are calculated directly from confusion probabilities, the prior

weights, and thus the model's performance, will vary depending on which confusion data are

used as input. Implicit in this design is the prediction that real language learners will learn

slightly differently depending on their own perceptual experience and language background. This

strikes me as a reasonable assumption.

For the model predictions reported above, I added together the confusion data listed in Table

2 and Table 3 from Wang and Bilger 1973 (henceforth WB). These data were chosen for several

130

reasons. First, WB's study included all of the crucial consonants tested in my study (i.e., [p, b, f, v, t, d, θ, ð]). Second, WB's participants were native English speakers, like my experimental participants. Third, the consonants were tested in CV syllables (WB:Table 2) and VC syllables (WB:Table 3). The target sounds in my experiments were located in a VCV context, so I added the data from the CV table and the VC table as a compromise (there was no VCV context collected in WB). Lastly, WB presented the stimuli in noise; the data from Table 2 and Table 3 represent the data across all of the signal-to-noise ratios (SNRs) that they tested (i.e., six SNRs ranging from −10 to +15).

Even though my experimental stimuli were presented without noise, I chose to use confusion data based on stimuli presented <u>with</u> noise for two reasons. First, the target sounds are all phonemes in English; thus, in clear speech, there are often too few confusions to reliably make assumptions about the relative similarity between sounds. For example, Singh and Black (1966) collected confusion data for English consonants in a CVC context without noise (i.e., the precise conditions of my experiments), but native English listeners made very few errors in that scenario (e.g., there were zero errors when the stimulus was [p] or [b]). Thus such data are not useful for generating the prior weights because they cannot differentiate pairs of sounds according to their similarity.

Second, under this theory, it is reasonable to assume that experimental participants use their overall experience as listeners throughout their lifetime as the basis of their P-map. In real life, people mostly hear speech in noisy environments. Only rarely do people hear speech that is comparable to clear, carefully pronounced laboratory speech. Thus, I would argue that confusion data in noise is more appropriate as the basis of the prior.

131

Despite these considerations, the reader may still be curious how dependent these results are on using any particular set of confusion data. To address these concerns, I ran several versions of the model with different confusion matrices as inputs for the prior. These models are summarized in Table 20. Note that confusion data from WB without noise are also included in this table. In that case, WB tested stimuli of different volumes (ranging from 20 dB to 115 dB) without noise. There were enough errors (due to the stimuli at lower volumes) to differentiate the sounds, so these matrices are included in Table 20.

As expected, model performance varies according to the precise confusion data used as the basis of the prior. Again, it is arguably a good property of the model that its predictions vary depending on the learner's perceptual "experience" (i.e., the confusion data used as the basis of the prior). Crucially, even as different confusion matrices are used to index perceptual similarity, the model's performance remains high. In particular, regardless of which confusion matrix is used, the biased model always outperforms the unbiased and high faith models: the lowest r^2 for the biased model is .77 compared to .25 for the unbiased model and .67 for the high faith model.

In sum, I conclude that although the precise predictions of the model change according to which confusion data are used to generate the prior, the overall success of the model is not dependent on using any particular confusion matrix.

Table 20. Performance of models (r^2) using different confusion data as the basis of the prior. The shaded lines represent the models reported above: the substantively biased model (top line) and the unbiased and high faith models (bottom lines). In all models, σ^2 is set to 0.6. WB = Wang & Bilger 1973; MN = Miller & Nicely 1955; C-etal = Cutler et al. 2004.

Source	Table #	Context	In noise?	r^2
WB 1973	2–3[33]	CV and VC	white noise	.95
WB 1973	2	CV	white noise	.93
WB 1973	3	VC	white noise	.92
WB 1973	6–7	CV and VC	none	.93
WB 1973	6	CV	none	.82
WB 1973	7	VC	none	.96
MN 1955	2–6	CV	white noise	.94
C-etal 2004	----[34]	CV and VC	babbled noise	.82
C-etal 2004	----	CV	babbled noise	.79
C-etal 2004	----	VC	babbled noise	.77
Unbiased model (for comparison)				.25
High faith model (for comparison)				.67

4.4.6 Learning the saltatory system

So far, we have established that human learners have a bias against saltations – they avoid them when faced with ambiguous data and they make errors when forced to learn explicit saltations. We have also seen that a MaxEnt learning model with a bias based on the P-map exhibits similar behavior when presented with the same training data. To gain a better understanding of how the model works, let us consider in greater detail how the constraint weights change throughout the learning process. For simplicity, I will consider only the *MAP constraints for labials, but the ones for coronals behave similarly.

The set of *MAP constraints begin (as a default) in a P-map-compliant orientation because their prior weights are calculated on the basis of perceptual confusability. For instance,

---

[33] Where multiple table numbers are given, the values were summed across those tables.

[34] The confusion matrices used from Cutler et al. (2004) are not taken from those reported in the paper, which only include the results for one of the SNRs that they tested. Instead, the data were taken from the supplemental webpage for their article, available at http://www.mpi.nl/world/persons/private/anne/materials.html. The matrices available at that webpage include data summed across all SNRs tested.

*MAP(p, v) has a hefty weight (3.65) because [p] and [v] are quite dissimilar, *MAP(p, b) has a medium weight (2.44), and *MAP(b, v) has a small weight (1.30) because [b] and [v] are very similar. These weights are given in Table 21 (left column).

Recall that in order to have a saltation, these constraints must eventually subvert the P-map-compliant hierarchy such that /p/ → [v] results in a smaller penalty relative to /b/ → [v]. Indeed, by taking a look at the weights after the model receives the relatively modest amount of training on explicit saltation from Experiment 2 (i.e., 18 p → v, 9 b → b, 18 t → ð, 9 d → d), we see that the weights have just made the switch (the relevant rows are shaded in Table 21). Due to the examples of /p/ → [v] in the training data, the weight of *MAP(p, v) begins to plummet (3.65 → 1.96). At the same time, the initially low weight of *MAP(b, v) is bolstered (1.30 → 2.02) by the examples of unchanging [b] in the data. Although *MAP(p, v) had a higher weight than *MAP(b, v) in the prior (3.65 vs. 1.30), their weights have now become quite similar (1.96 vs. 2.02) based on evidence from the training data. The weight of *MAP(p, b) is also increased (2.44 → 2.94) to ensure that /p/ changes all the way to [v], rather than "getting stuck" at intermediate [b].

At the stage of learning that is based only ↘on the training data from Experiment 2, the grammar is clearly in a transitional state, and this transitional state illustrates why the learner exhibits the anti-saltation learning bias that we see with experimental participants. The training data in Experiment 2, though few in number, are entirely consistent with a saltatory pattern. However, because the model must work against a prior that biases it towards non-saltatory outcomes, the limited amount of training data encountered by the model is not sufficient to completely subvert the constraint hierarchy implicit in the P-map prior. Thus, even though there

is no variation in the training data, the ongoing influence of the prior leads the model to predict

errors that are comparable to those made by human learners in the experiment.

Table 21. Weights for the markedness constraints and the *MAP constraints (labials only) over the course of learning.

Constraint	Prior weight	After training data from Experiment 2	After 1000 training data of each type
*V[−voice]V	0	2.45	4.82
*V[−cont]V	0	1.05	0.30
*MAP(p, v)	3.65	1.96	0
*MAP(b, v)	1.30	2.02	4.75
*MAP(p, b)	2.44	2.94	4.62
*MAP(p, f)	1.34	1.74	1.75
*MAP(b, f)	1.96	2.02	2.29
*MAP(f, v)	2.56	2.56	2.56

Moving beyond this initial learning bias, we need to ensure that the model can eventually

learn a saltatory system. Saltations are attested in real languages (see section 2.2), so it must be

possible for children to successfully learn such a system, even if it is initially dispreferred. To

test that the model can properly reach this final state (i.e., it can learn complete, categorical

saltation), I trained the model with the following input: 1000 cases of [p] → [v], 1000 cases of [t]

→ [ð], 1000 cases of unchanging [b], and 1000 cases of unchanging [d]. Thus, the type of input

was similar to Experiment 2, but the amount of training data was much more extensive,

representing the large amount of input a real child would receive.

The resulting constraint weights are given in the right column of Table 21. The weights of

the *MAP constraints continue to move in the same direction. With this much training data, the

relevant constraints manage to completely subvert the default hierarchy imposed by the P-map:

*MAP(p, v) reaches a weight of 0 (i.e,. it has no effect on the outcome) whereas *MAP(b, v) and

*MAP(p, b) reach substantial weights of 4.75 and 4.62, respectively. The repeated onslaught of

input that goes against the prior leads the model to gradually overcome the prior's influence on its quest to successfully account for the observed data.

Ideally, the resulting model should predict [p] → [v] and [t] → [ð] near 100% of the time, and also that [b] and [d] remain unchanged near 100% of the time. The actual predicted results are given in Table 22.

Indeed, the model predicts that the saltatory system will be learned essentially perfectly, with virtually equal predictions for the labials and coronals. Only 0.6% of the time does the model predict a mistake on the saltatory change, which is low enough to be due to occasional speech errors.[35] Likewise, the model predicts that the intermediate sounds will be changed in error only 1.2% of the time for labials and 1.1% of the time for coronals. Again, this is low enough that such errors could be considered speech errors. Bolognesi (1998, p. 36) indeed reports that native speakers occasionally spirantize intervocalic voiced stops (in error), but that such errors occur only rarely.

Table 22. Model predictions when provided with 1000 cases of each observation during training, showing that the model can learn a saltatory system.

Labials		Coronals	
Outcome	Prediction (in %)	Outcome	Prediction (in %)
p → v	99.4	t → ð	99.4
p → p	0.6	t → t	0.6
b → v	1.2	d → ð	1.1
b → b	98.8	d → d	98.9

---

[35] With even more training data, this percentage would get even lower, but never down to 0%. It is not possible for an output to have a prediction that is truly 0% in MaxEnt; it can, however, reach such a low number that the predicted probability is practically 0% (i.e., so low that the output might never occur in a lifetime).

4.5 A harder test: The production study

So far, I have used the results of a two-alternative forced choice (2AFC) task both to demonstrate that adults have an anti-saltation bias (Chapter 3) and as the basis of comparison for model predictions (current chapter). In this section, I present the results of a production study designed to replicate Experiment 1 (section 3.4). The production task provides a more difficult task for the experimental subjects because they have to offer their own responses without the benefit of hearing response options. It also serves as a more difficult test of the model, which is capable of predicting probabilities for several output candidates at once.

The overall design of the study is similar to Experiment 1 as described in section 3.4, with certain modifications adapted from a production experiment by Skoruppa et al. (2011). The experiment still had three phases: exposure, verification, and generalization. In the exposure phase, participants heard a singular word (e.g., [falap]) paired with a picture of a singular item. Instead of hearing a plural word, they were asked to guess the plural word out loud upon seeing a plural picture of the item. After making a guess, they received feedback by hearing the correct answer over headphones. At first, participants had no way to know the correct answer, but as training continued, they would gradually learn the correct patterns. This more active method of training (i.e., requiring a spoken response and giving immediate feedback) was chosen over the more passive training used in Experiment 1 with the hope that it would result in faster, more effective learning.

Participants were divided into two conditions: a Potentially Saltatory condition and a Control condition. Like in Experiment 1, participants in the Potentially Saltatory condition learned alternations between dissimilar sounds (i.e., [p ~ v] and [t ~ ð]) during exposure, but had no examples of intermediate [b, d, f, θ] during training. Participants in the Control condition

instead learned alternations between similar sounds (i.e., [b ~ v] and [d ~ ð]) during exposure, and had no examples of [p, t, f, θ]. Note that in the Control condition, these untrained sounds were not intermediate between the alternating sounds.

After the exposure phase, participants completed the verification phase, which included words that they had already encountered during exposure. The task was the same as in exposure, except that they no longer received feedback. Because the task was a production task and coding the responses online was not practical, it was no longer possible to require participants to reach an accuracy criterion to move on to the generalization phase. Instead, participants' responses were recorded and coded offline. Data from participants who did not achieve an overall accuracy of 80% (on alternating sounds and filler sounds combined) in the verification phase were not included in the analysis because they had not properly learned the alternations during exposure.

In the generalization phase, participants completed the same task but with novel words, including words ending in target sounds not presented during exposure. Participants were predicted to exhibit the same effect found in the 2AFC task: they should change untrained intermediate sounds (Potentially Saltatory condition) more frequently than comparable untrained sounds that are not intermediate (Control condition). Changing untrained sounds in the Potentially Saltatory condition allows the participants to avoid a saltatory system, but doing so in the Control condition does not.

4.5.1 Method

4.5.1.1 Participants

Eighty-six undergraduate students in introductory psychology or linguistics classes at UCLA completed the experiment for partial course credit. The participants were randomly assigned to either the Potentially Saltatory condition or the Control condition. None of the participants had participated in Experiments 1 or 2 from Chapter 3.

4.5.1.2 Materials

Exposure phase. For the exposure phase, 48 nonwords of the form CVCVC (e.g., [kamap]) were used as singular stimuli for the Potentially Saltatory condition. Many of these forms were in fact taken from the nonwords used in Experiment 1. Half of the nonwords ended in the target sounds {p, t}, 12 of each, and half of the nonwords ended in one of the filler sounds {n, l, ɹ}, 8 of each. The initial consonant sounds were drawn from the set {p, b, t, d, k, g, f, θ, s, ʃ, m, n, l, ɹ}. The medial consonants were chosen from the more limited set of filler sounds {m, n, l, ɹ} there would not be unintentional distributional information relevant to the target sounds. Vowels were drawn from the set {i, a, u}. In all other ways, nonwords were created in the same way described for Experiment 1.

For each of the 48 singular nonwords, a plural form was also created by adding the vowel [i] to the end of the singular nonword. For singular items ending in filler sounds, there was no change in the final consonant for the plural form (e.g., singular [luman], plural [lumani]). For nonwords ending in {p, t}, the final consonant was changed to the corresponding voiced fricative, either [v] or [ð] (e.g., singular [kamap], plural [kamavi]). Stress was placed on the final vowel of singular items and the penultimate vowel of the plural forms.

139

For the singular nonwords ending in {p, t}, corresponding nonwords for the Control condition were created by changing each final [p] to [b] and each final [t] to [d], as in Experiment 1. Except for this modification, the lists of nonwords used in the Potentially Saltatory condition and the Control condition were identical. For instance, singular [kamap] and plural forms [kamapi] and [kamavi] in the Potentially Saltatory condition corresponded to singular [kamab] and plural forms [kamabi] and [kamavi] in the Control condition.

Each set of nonwords was matched with the same pairs of singular and plural images from Experiment 1. The pictures were made up of clipart-style images or small photographs of everyday nouns taken from the Internet.

Verification phase. For the Potentially Saltatory condition, 24 of the singular nonwords (6 *p*-final, 6 *t*-final, and 12 fillers), along with their associated pictures, were chosen from the set of nonwords in the exposure phase for use in the verification phase. In the Control condition, the corresponding set of nonwords was used.

Generalization phase. For the generalization phase, 76 new singular nonwords were created in the same manner described above (many of them taken from Experiment 1). For the Potentially Saltatory condition, 24 ended in {p, t} (12 of each), 28 ended in the filler sounds (6 each of {n, l, ɹ}, 10 of {m}), and 24 ended in the intermediate sounds {b, d, f, θ} (6 of each). For the Control condition, the same set of words were used except word-final [p] was changed to [b], word-final [t] was changed to [d], and vice versa.

Stimuli recording and experimental apparatus. The stimuli were recorded in the same way as in Experiments 1 and 2 (see section 3.4.1.2), by the same speakers but on a different occasion.

The experiment was conducted in a quiet room on a Dell computer equipped with a 20-inch monitor and Sony MDR-V200 headphones. The experimental software E-prime (version 2.0) was used to present the stimuli. To record their verbal responses during the verification and generalization phases, participants wore, in addition to their headphones, a Shure SM10A head-mounted microphone, whose signal ran through an XAudioBox pre-amplifier and A-D device. The recordings were done using PcQuirerX at a sampling rate of 22,050 Hz.

4.5.1.3 Procedure

The experiment consisted of three phases: exposure, verification, and generalization. Before the first exposure trial, participants completed three practice trials with the experimenter in the room to ensure that they understood the task. The practice trials consisted of three additional filler items.

In the exposure phase, participants were instructed that they would be learning words in a foreign language. Participants completed 48 unique trials in this phase. Each trial began with a picture showing a singular object appearing in the center-left part of the computer screen. After the picture had been displayed for one second, the singular nonword for that item was played over headphones. The singular picture disappeared 2.5 seconds after the sound file began playing, and the corresponding plural picture immediately appeared in the center-right part of the screen, along with a row of question marks just under it. At this point, participants provided a guess for the plural picture out loud. After making their guess, they pressed the space bar. The row of question marks changed to "The correct answer is..." and the participants heard the

141

correct plural forms for that trial over headphones. After hearing, the correct plural word, the participants pressed the spacebar to move on to the next trial. Nonwords were only presented in auditory form, never in orthography. The order of trials in this phase, as well as in the following two phases, was randomized anew for each participant by E-prime. The exposure phase lasted approximately 15 minutes.

In the verification phase that followed, participants were tested on 24 words that they had heard during the exposure phase. The trials were identical to those in the exposure phase except participants no longer received feedback. After providing a response out loud for the plural item and pressing the spacebar, the next trial began immediately. This phase lasted approximately 10 minutes.

After the 24 verification trials, participants moved into the generalization phase, where they were tested on 76 novel words, including words ending in untrained target sounds. Trials in this phase were otherwise identical to those in the verification phase. The generalization phase lasted approximately 15 minutes.

4.5.2 Coding and exclusions

Responses in the generalization phase were coded offline by an English speaker with phonetics training who was familiar with the purpose of the study (the author). Responses in the verification phase were coded as correct or incorrect. For the generalization phase, the coder transcribed the final consonant of the plural word given as a response. To check for reliability, a second coder, who was an English speaker with some phonetic training (an introductory course) but was not familiar with the purpose of the study, also coded each response. The coders agreed on 95.4% of the trials.

Participants who did not have an overall accuracy of 80% in the verification phase were not included in the analysis of the generalization phase. Thirty-eight participants did not reach criterion (15 out of 33 from the Potentially Saltatory condition, 23 out of 46 from the Control condition). That left a total of 41 participants in the analysis, 18 in the Potentially Saltatory condition and 23 in the Control condition.

4.5.3 Experimental results

To determine if the production study replicated the basic saltation avoidance effect found for the untrained sounds in the 2AFC study, I first consider only how often participants changed untrained sounds to voiced fricatives or left them unchanged (i.e., the two options from the 2AFC task). Only responses in the generalization phase were included in the analysis. Overall, it appears that there was a greater tendency to change the untrained sounds to voiced fricatives in the Potentially Saltatory condition than in the Control condition, both for the untrained stops (33.3% changed vs. 16.3% changed) and the untrained fricatives (29.2% changed vs. 12.3% changed), consistent with the results from the 2AFC task in Experiment 1.

To assess these differences, the trials in which one of these two options were chosen were analyzed using a mixed effects logistic regression model (see Jaegar, 2008), the same statistical analysis used in the 2AFC (see section 3.4.2). The model predicted the log odds of a changing response. The model was implemented in R (R Core Development Team, 2008) using the *lme4* package (Bates, Maechler, & Dai, 2008). To compare models in a subset relation, likelihood ratio tests were conducted using the anova() function (Baayen, Davidson, & Bates, 2008).

The model contained fixed effects for Condition (Potentially Saltatory vs. Control), Sound Type (untrained stops vs. untrained fricatives), and a Condition x Sound Type interaction.

Random intercepts were included for subjects and words. By-subject random slopes for Sound Type were also included because they significantly improved mode fit according to a likelihood ratio test, $\chi_2(2) = 24.44$, $p < .001$.

The fixed effects from the resulting model are provided in Table 23. The significant negative intercept indicates that the untrained target sounds were changed to voiced fricatives infrequently overall in the Control condition (which acts as the baseline in this model). Crucially, the significant effect of Condition indicates that the untrained target sounds were changed to voiced fricatives significantly more often in the Potentially Saltatory condition (where they were intermediate) than in the Control condition (where they were not intermediate); thus the overall saltation avoidance effect observed in the 2AFC was replicated in the production task.

The effects of Sound Type and the Condition x Sound Type interaction were not significant in the model (unlike in Experiment 1, see section 3.4.2.2), and likelihood ratio tests indicate that neither the interaction term ($\chi^2(1) = 1.19$, $p = .28$), nor the factor of Sound Type ($\chi^2(2) = .002$, $p = .96$), significantly improved model fit compared to models without those terms.

Table 23. Summary of the fixed effects for the production study.

Predictor	Estimate	Standard error	z-value	p-value
Intercept	−2.76	.47	−5.84	<.001
Condition = *Potentially Saltatory*	2.09	.63	3.32	<.001
Sound Type = *Untrained stops*	.51	.52	.98	.33
Interaction = *Potentially Saltatory & Untrained stops*	− .81	.69	−1.18	.24

4.5.4 Comparison with model predictions

The results of the production experiment for the Potentially Saltatory condition and the Control condition are presented in Table 24 and Table 25, respectively, along with the

144

predictions of the substantively biased model. These model predictions come from the same

grammar used in section 4.4; the only difference is that the model considered four possible

outputs for each input at test instead of being limited to the two from the 2AFC task. The points

of greatest divergence between the model predictions and the experimental results have been

shaded in the tables for easy comparison.

Comparing the model predictions to the experimental results, there are three key areas

where the experimental results seem to diverge substantially from the predictions: (1) the model

overestimated the amount of generalization in the Potentially Saltatory condition (Table 24), (2)

participants performed better than the model predicted on the alternations that they were trained

on in both conditions, and (3) the model predicts massive feature-based generalization in the

Control condition (i.e., parallel generalization from [b] → [v] to [p] → [f], and similarly for the

coronals) whereas there was very little generalization of this sort in the actual experiment. I

consider each of these three issues in turn.

Table 24. Experimental results (production) and model predictions for the Potentially Saltatory condition.

	Experimental results					Model predictions				
	Response					Predicted response				
input	p	b	f	v	other	p	b	f	v	other
p	1.9	1.9	0.4	94.5	1.4	10.9	6.2	10.4	72.5	--
b	0.9	56.5	0	40.7	1.9	0.2	35.0	3.5	61.2	--
f	0.9	0	64.8	32.4	1.9	1.2	10.3	52.1	36.4	--
	t	d	θ	ð	other	t	d	θ	ð	other
t	1.4	2.8	3.7	91.7	0.5	12.0	5.3	7.4	75.3	--
d	0	73.2	0	25.9	0.9	0.2	37.8	2.2	59.8	--
θ	9.3	0	31.0	25.9	30.0	0.6	4.7	40.5	54.2	--

Table 25. Experimental results (production) and model predictions for the Control condition.

	Experimental results					Model predictions				
	Response					Predicted response				
input	b	p	f	v	other	b	p	f	v	other
b	6.5	0	0.4	92.4	0.7	15.2	0.7	5.4	78.7	--
p	1.5	68.9	6.5	18.8	4.3	4.9	33.3	52.5	9.2	--
f	0	0	82.6	14.5	3.6	2.5	3.6	82.6	11.3	--
	d	t	θ	ð	other	d	t	θ	ð	other
d	8.7	0.4	6.2	77.5	7.2	16.8	0.6	3.9	78.7	--
t	2.2	73.9	8.7	13.8	1.5	4.9	43.8	38.0	13.3	--
θ	3.6	1.5	57.2	10.1	23.2	1.5	1.8	76.6	20.1	--

4.5.4.1 Overgeneralization to intermediate untrained sounds

Let us first consider the issue of overgeneralization in the Potentially Saltatory condition. In the 2AFC task, generalization to intermediate sounds in the Potentially Saltatory condition was high (67–70% for voiced stops, 41–49% for voiceless fricatives). The substantively biased model was successful at capturing this degree of generalization (see section 4.4.1). In the production task, however, the amount of generalization was much lower (26–41% for voiced stops, 26–32% for voiceless fricatives), but the model still predicts that generalization will be high (Table 24).

Intuitively, this seems to be a task-specific effect: learners are less willing to posit new alternations in the production task than they are in the 2AFC, where they are given the possible response options. Perhaps novel alternations are more appealing if they have been presented as one of two response options. To account for this difference in tasks, I propose that learners have access to an additional constraint *ALTERNATE which can be "turned up" in cases where learners have a reason to be more conservative, such as in production tasks. It penalizes any alternation whatsoever; essentially, it acts as an general preference for uniform paradigms (e.g., Hayes, 1997; Steriade, 2000).

146

I added *ALTERNATE into the model with a prior μ of 1 and a σ^2 of .0001, meaning that the weight of *ALTERNATE is essentially forced to remain at 1, given the amount of training data available in these experiments. Without a low σ^2, the weight of *ALTERNATE would drop rapidly in response to the cases of alternation in the training data to a point where it would have little effect on the outcome for untrained cases. Instead, I wanted *ALTERNATE to act as a general bias that can be scaled according to task, but is not subject to (rapid) learning. The model was retrained, and the resulting predictions are presented in Table 26 and Table 27. The cases where the model was previously overgeneralizing are bolded instead of shaded.

Table 26. Experimental results (production) and model predictions for the Potentially Saltatory condition, with a *ALTERNATE constraint.

	Experimental results					Model predictions				
	Response					Predicted response				
input	p	b	f	v	other	p	b	f	v	other
p	1.9	1.9	0.4	94.5	1.4	15.4	4.6	8.3	71.7	--
b	0.9	56.5	0	**40.7**	1.9	0.1	51.9	2.0	**46.0**	--
f	0.9	0	64.8	**32.4**	1.9	0.5	4.9	70.6	**24.0**	--
	t	d	θ	ð	other	t	d	θ	ð	other
t	1.4	2.8	3.7	91.7	0.5	16.5	3.8	5.7	74.0	--
d	0	73.2	0	**25.9**	0.9	0.1	54.8	1.2	**43.9**	--
θ	9.3	0	31.0	**25.9**	30.0	0.3	2.4	59.0	**38.3**	--

Table 27. Experimental results (production) and model predictions for the Control condition, with a *ALTERNATE constraint.

	Experimental results					Model predictions				
	Response					Predicted response				
input	p	b	f	v	other	p	b	f	v	other
b	6.5	0	0.4	92.4	0.7	20.3	0.4	4.9	74.3	--
p	1.5	68.9	6.5	18.8	4.3	2.4	46.3	44.2	7.1	--
f	0	0	82.6	14.5	3.6	0.6	0.9	94.1	4.3	--
	d	t	θ	ð	other	d	t	θ	ð	other
d	8.7	0.4	6.2	77.5	7.2	21.4	0.3	3.3	74.9	--
t	2.2	73.9	8.7	13.8	1.5	2.2	57.7	30.3	9.7	--
θ	3.6	1.5	57.2	10.1	23.2	0.4	0.5	91.1	8.1	--

The amount of generalization to intermediate sounds predicted in the Potentially Saltatory condition is now much more reasonable, especially for the labials. The model still overgeneralizes for coronals, but the amount of overgeneralization has decreased.[36] One might expect that *ALTERNATE would also cause the model to predict a higher percentage of faithful outputs for the sounds that it was trained should alternate, which would be problematic because the model was already predicting too little alternation in those cases (see below). But in fact, the effect on the trained alternations is very small. A glance at the weights, shown in Table 28, reveals why this is so. For simplicity, only *MAP constraints for labials are shown.

In response to the addition of *ALTERNATE, three constraints affecting input [p] are adjusted so that [p] will continue to alternate: both markedness constraints are adjusted upwards and *MAP(p, v) is adjusted slightly downward. By contrast, only one constraint affecting each of the untrained intermediate sounds is adjusted, which is the relevant markedness constraint (i.e., *V[–voice]V for [f] and *V[–cont]V for [b]). The faithfulness constraints for the untrained sounds cannot be adjusted because there is not training data for those sounds. As a result, the addition of *ALTERNATE has a large effect on unseen alternations, but only a small effect on trained alternations. Overall, adding *ALTERNATE improves the model's fit to the experimental data from the production task ($r^2 = .78$ without *ALTERNATE, .87 with *ALTERNATE).[37]

---

[36] The smaller amount of generalization for the coronals could be due to a separate bias against [ð], which is phonetically marked and rare in English (in terms of word types, not tokens; Thatte, 2011) and might thus be marginally dispreferred in the production of new words. In support of this possibility, I note that [ð] is frequently turned into stops by speakers in normal speech. I will not attempt to model that possibility here.

[37] When calculating the model's fit to the production data, I also redistributed the cases of input [θ] coded as "other" proportionally between the two leading response options for that row (i.e., [θ] and [ð]). Almost all of the (substantial number of) "other" cases were in fact responses of [f] or [v], indicating that the [θ] had been misperceived as [f]. Because the model never misperceives and has no "other" output category, it could not possibly match the observed results in cases with so many "other" responses. This modification resulted in a modest improvement in the model's fit, increasing r^2 from .82 to .87.

Table 28. Behavior of weights in the Potentially Saltatory condition with and without *ALTERNATE.

Constraint	Prior weight	After training, without *ALTERNATE	After training, with *ALTERNATE
*V[–voice]V	0	2.20	2.48
*V[–cont]V	0	1.86	2.18
*MAP(p, v)	3.65	2.17	2.12
*MAP(b, v)	1.30	1.30	1.30
*MAP(p, b)	2.44	2.77	2.69
*MAP(p, f)	1.34	1.90	1.79
*MAP(b, f)	1.96	1.96	1.96
*MAP(f, v)	2.56	2.56	2.56
*ALTERNATE	1.00	-----	1.00

4.5.4.2 Underperformance on trained alternations

A second problem of the model is that it predicts too few correct responses for the trained alternations in both the Potentially Saltatory and the Control conditions. In the actual experiment, participants had an accuracy above 90% on all of the trained alternations with the exception of [d] → [ð] (where, again, a dispreference for [ð] may have played a role). This discrepancy may in fact be an artifact of the exclusion criteria. Only participants who were 80% accurate in the verification phase (of which half of the trials consisted of the trained alternations) were included in the analysis of the generalization phase. This aspect of the experimental design may have artificially increased the accuracy on the trained alternations in the generalization phase.

4.5.4.3 Feature-based generalization

The final discrepancy between the model predictions and the experimental results stems from the model's prediction of feature-based generalization in the Control condition. The model predicts, for example, that learning [b] → [v] will lead to a substantial amount of generalization to [p] → [f]. Even when *ALTERNATE is included in the model, it predicts that the response will

be [f] in 44.2% of trials with input [p] (see Table 27). This is not a crazy prediction because it follows natural classes; the rule can be stated very simply as "stops becomes fricatives between vowels." The model predicts this type of generalization due to its feature-based markedness constraints. Specifically, the weight of *V[–cont]V must be increased to motivate the [b ~ v] alternation, but doing so also motivates a change from [p] to [f]. This was not an issue in the 2AFC task, where [f] was not a possible response option for input [p].

In the experimental task, participants occasionally generalized according to the feature [continuant], but the amount of generalization was extremely modest (6.5% [p] → [f], 8.7% [t] → [θ]; see Table 27). This appears to be a case where the model is legitimately making the wrong prediction, which raises serious questions about the role of feature-based generalization in phonological learning. The concept of generalization on the basis of features or natural classes has long played in important role in phonology, back to *SPE* (Chomsky & Halle, 1968) and before. However, it remains unclear if this type of parallel generalization plays a role during learning, and if so, when and how it plays a role. Why do we see the saltation avoidance effect (Experiments 1 and 2), which under this analysis is due in part to featural generalization, but we do not see the type of parallel generalization (e.g., from b → v to p → f) also predicted by the model for the production study? Given these issues, an in-depth investigation of the role of generalization in phonological learning should, I believe, be an important goal of future experimental and computational work in phonology.

4.5.5 Overall model performance

Figure 12 shows the predictions of the substantively biased model (with *ALTERNATE) plotted against the aggregate experimental results from the production experiment. We can see

that the fit between the model predictions and the experimental results is not quite as good as in

the 2AFC task (see section 4.4.3), but the fit is still very good ($r^2 = .87$).

Figure 12. Predictions of the substantively biased model plotted against the experimental results from the production experiment. Overall $r^2 = .87$.
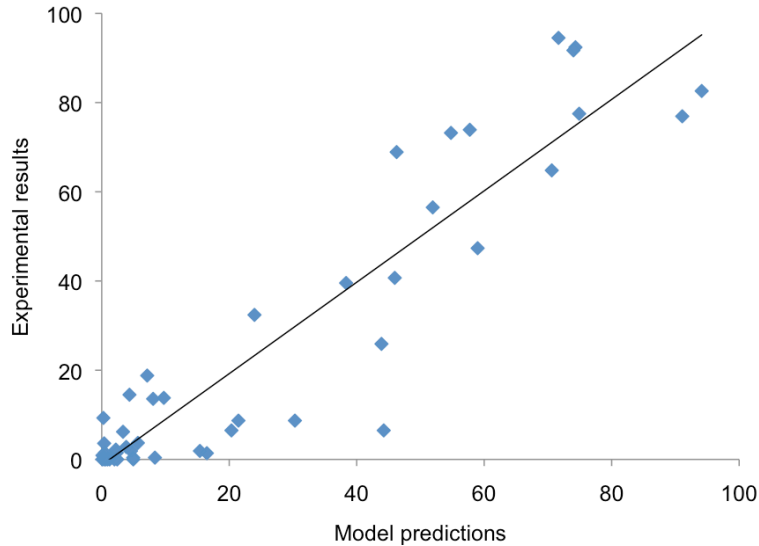
For comparison, the predictions of the unbiased model are plotted against the experimental

results in Figure 13. For the best comparison, the unbiased model was also augmented with a

*ALTERNATE constraint with the same prior values that it had in the biased model (i.e., μ of 1 and

a σ^2 of .0001). As the plot shows, the unbiased model is much less successful at predicting the

experimental results; as a result, the fit between the model predictions and the experimental

results is much lower ($r^2 = .39$).

Figure 13. Predictions of the unbiased model plotted against the experimental results from the production experiment. Overall $r^2 = .39$.
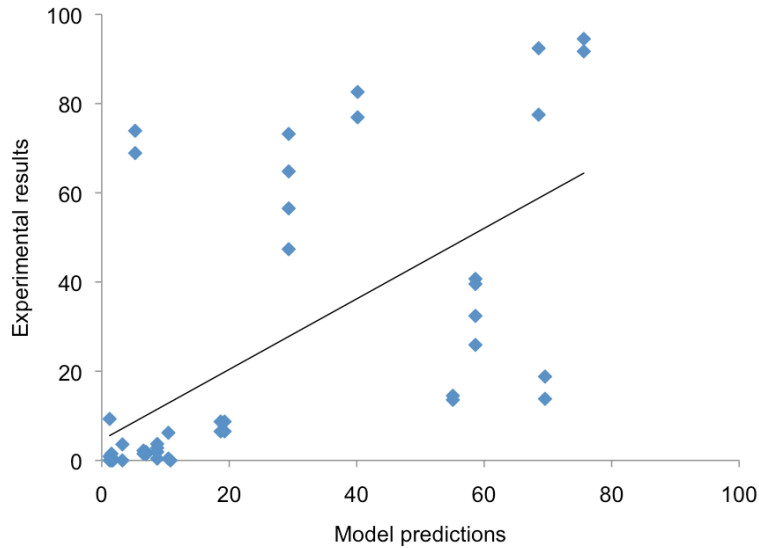
Finally, I also compared the predictions of the high faith model to the substantively biased model. Once again, the high faith model was also equipped with *ALTERNATE for the fairest comparison. The model's predictions are plotted against the experimental results in Figure 14. The high faith model performs much better than the unbiased model ($r^2 = .82$ vs. .39), indicating that having non-zero default weights for the *MAP constraints is important for the model. The substantively biased model, in turn, outperforms the high faith model ($r^2 = .87$ vs. .82) indicating that the benefit of the substantive bias extends beyond the benefit associated just with having non-zero weights for the *MAP constraints.

Figure 14. Predictions of the high faith model plotted against the experimental results from the production experiment. Overall $r^2 = .82$.
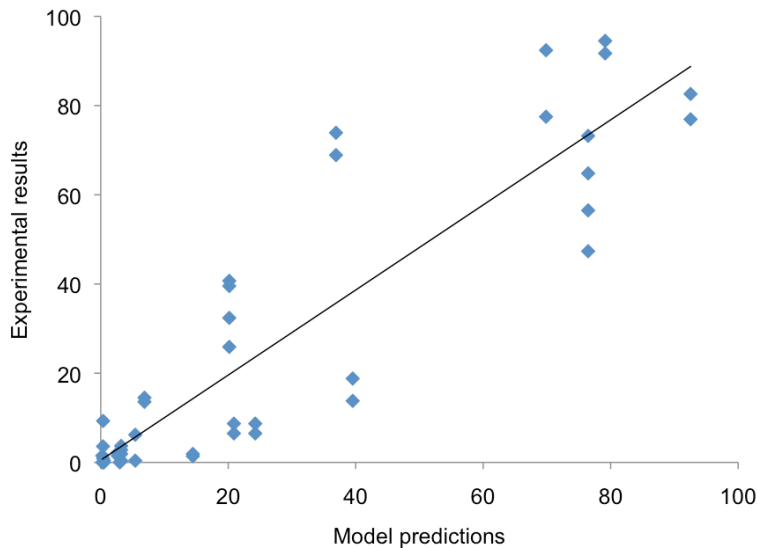
In terms of r^2, the increase from .82 to .87 due to the substantive bias is fairly small. However, a qualitative evaluation of the plots suggests that the substantive bias is making a meaningful contribution. The high faith model (Figure 14) is successful at predicting the high and low values, but it does not perform well at predicting the moderate values (i.e., there are no data points on the middle portion of the regression line in Figure 14) because it fails to properly differentiate many of the data points (notice the vertical lines of data points in Figure 14). By contrast, the substantively biased model (Figure 12) is successful at predicting moderate values – points fall all along the regression line, including the middle portion.

One potential reason for the small difference in r^2 is the large number of near-zero values being modeled in the production study (most potential responses were chosen infrequently). Both models perform well on the high and low values, and r^2 is highly sensitive to extreme values. Indeed, considering only the 16 experimental values falling between 10% changed and 80% changed, the substantively biased model has a much better fit to the experimental results

153

compared to the high faith model (r^2 = .80 for the substantively biased model; r^2 = .51 for the high faith model).

In sum, the difference in r^2 between the substantively biased model and the high faith model is not substantial for the production experiment overall. But the substantive bias appears to play a meaningful role in differentiating the moderate values, which are the cases where small differences in perceptual similarity are most likely to make a difference.

4.5.6 Summary of production study

We have seen that the basic saltation avoidance effect observed in the 2AFC task was replicated in a production task. Comparing the artificial language learning tasks, the results also suggested that participants are more conservative in production tasks than in 2AFC tasks. I proposed that one way to model this difference in conservativeness was to include a general *ALTERNATE constraint that penalizes any alternations, with a prior weight that varies depending on task. Finally, we saw the substantively biased model outperforms both the unbiased model and the high faith model (especially for the moderate values) when it comes to predicting the responses in the production task, suggesting that perceptual similarity played a role in the participants' learning.

4.6 General discussion

4.6.1 Summary of chapter

To summarize, we have seen that the case of saltation is problematic for traditional phonological frameworks. First, saltation is attested in real languages, so it must be possible for

children to learn a saltatory system. Any theory that cannot generate saltations, such as classical OT, thus cannot account for the existence for these patterns.

Second, artificial grammar experiments with adults have demonstrated two separate, but related, cases where learners acquire novel alternations in a biased way – in particular, in a way that disfavors saltation. Experiment 1 showed that learners generalize when learning novel alternations, but only when doing so would avoid a saltation. In other words, they will generalize from alternations between dissimilar sounds to more similar sounds (in particular, to intermediate sounds), but not vice versa. Experiment 2 showed that adults have difficulty learning saltation even when it is explicitly presented in the training data.

In light of these facts, I proposed a phonological framework that can adequately account for both the possibility of learning saltations as well as the dispreference for them. The model had three main components that played a role in its success:

1) ***MAP constraints:** Zuraw (2007)'s family of faithfulness constraints crucially made it <u>possible</u> for saltations to be learned by making it possible for correspondences between dissimilar sounds to be preferred over correspondences between similar sounds.

2) **P-map bias:** A learning bias based on Steriade's (2001/2008) P-map and her principle of minimal modifcation was responsible for the saltation avoidance observed in experiments. In particular, it explains why learners would generalize alternations to intermediate sounds, but not to other nearby sounds that are not intermediate. It also accounts for why learners erroneously change intermediate sounds when learning explicitly saltatory alternations.

155

3) **<u>MaxEnt learning</u>:** The architecture of the MaxEnt learning model was the final component that pulled everything together. The prior term served as a perfect vehicle for implementing the P-map bias computationally. From there, the learning process itself is the reason why we observe a bias initially (due to the prior), but also reach the final state of (effectively) categorical saltation after sufficient amounts of training data. MaxEnt allows the prior to be overturned <u>gradually</u> through learning, much in the same way that, we can hypothesize, the child might learn saltation.

The predictions of the substantively biased model provided an excellent fit to the experimental data, both in the 2AFC tasks (from Chapter 3) and in the production task replication (current chapter). In particular, the substantively biased model outperformed a completely unbiased model (prior weight of 0 for all constraints) as well as a model in which the *MAP constraints have non-zero weights, but do not differ based on perceptual similarity. These findings together support the view a substantive bias based on perceptual similarity plays a role in the learning of phonological alternations.

In the remainder of this chapter, I will consider two outstanding issues with respect to the learning model: how it compares with Wilson's (2006) model and how it relates to discussions of the initial state of the grammar.

4.6.2 Comparison with Wilson's (2006) implementation

The general approach to biased phonological learning taken here follows the approach taken by Wilson (2006): use the prior of the MaxEnt model to implement a substantive bias. However, in the details of the implementation, our approaches are actually quite different.

Wilson was interested in predicting different rates of velar palatalization (i.e., /k/ → [tʃ] and /g/ → [dʒ]) depending on whether the following vowel was [i], [e], or [a]. Perceptually, [k] and [tʃ] are most similar before [i], followed by [e], and least similar before [a]. All else being equal, [d] and [dʒ] are less perceptually similar than [k] and [tʃ]. Typological observations are consistent with the predictions of the P-map: velar palatalization is most common before high vowels and least common before low vowels.

In contrast with my model, Wilson implements the substantive bias by setting different σ^2 values for the various markedness constraints he uses to drive palatalization. Recall that in my model, the substantive bias was instead implemented by setting a different μ for each faithfulness constraint (i.e., *MAP constraint).

Wilson included 12 different markedness constraints in his model. The markedness constraints motivate palatalization by penalizing sequences of [k] or [g] followed by a specific vowel (i.e., *ki, *ke, *ka, *gi, *ge, *ga) or a general class of vowels (i.e., *kV$_{[-low]}$, *kV$_{[-high]}$, *kV, *gV$_{[-low]}$, *gV$_{[-high]}$, *gV). The μ for each of these constraints was set at 0. The σ^2 for each constraint was calculated based on the perceptual similarity between the penalized input consonant and the palatalized output consonant that would result; for example, the σ^2 for *ki was calculated based on the perceptual similarity of [k] and [tʃ] before [i].[38] The resulting set of σ^2 values affected how easily the weight for each markedness constraint could be moved from 0. For instance, *ki received a relatively high σ^2 whereas *ka received a lower σ^2 because [k] and

---

[38] Perceptual similarity was calculated by Wilson using the generalized context model of classification (GCM; Nosofsky, 1986), taking into account featural similarity, acoustic similarity (peak spectral frequency), confusability (based on confusion data from Guion, 1998), and overall response bias. See Wilson (2006) for a detailed account.

[tʃ] are more similar before [i] than before [a]. As a result, the weight of *ki would be able rise

quickly in the face of training data relative to the weight of *ka, which would ultimately result in

a greater tendency to palatalize /ki/ compared to /ka/. The model also contained two faithfulness

constraints, one penalizing changes to /k/ and one penalizing changes to /g/, which were assigned

high (but otherwise fairly arbitrary) values for μ and σ^2.

An important issue for modeling Wilson's experimental results was predicting

generalization (e.g., that participants would generalize palatalization from the mid vowel context

to the high vowel context but not vice versa). Generalization in the model was due to the set of

markedness constraints targeting [k] or [g] in general contexts, such as *kV$_{[-low]}$. With a

Gaussian prior, MaxEnt models prefer to spread responsibility between several constraints rather

than putting all of the weight onto a single constraint. For instance, cases of /ke/ → [tʃe] would

boost the weight of the general constraint *kV$_{[-low]}$ in addition to the more specific constraint

*ke. As a result, rates of /ki/ → [tʃi] would be increased at test even if no examples of /ki/

appeared during training.

It is worth noting that Wilson's implementation of the substantive bias as a property of the

markedness constraints is somewhat odd conceptually. Traditionally, markedness constraints are

only allowed to evaluate each output candidate's surface characteristics on an individual basis.

In Wilson's model, the markedness constraints must have access to the perceptual relationship

between the faithful candidate (meaning it must first know which candidate is faithful) and one

of the competitor candidates (c.f. targeted constraints; Wilson, 2001). By putting the bias on the

faithfulness side, as in my model, this perceptual relationship is considered by constraints that

are already assumed to evaluate the correspondence between two segments.

I ran a separate model to determine whether my implementation can also account for

Wilson's results. The constraint set used in my model is shown in (36):

(36) *Constraints used in the reanalysis of Wilson (2006)*

    a. Markedness constraints

        *kV
        *gV

    b. Faithfulness constraints

        *MAP(k, tʃ)/_i
        *MAP(k, tʃ)/_e
        *MAP(k, tʃ)/_a
        *MAP(g, dʒ)/_i
        *MAP(g, dʒ)/_e
        *MAP(g, dʒ)/_a
        *ALTERNATE

My model contained only two markedness constraints to motivate palatalization, *kV and *gV,

which penalize [k] or [g], respectively, when they occur before any vowel. It also contained a set

of *MAP constraints banning palatalization in the phonological contexts relevant for Wilson's

experiment (i.e., before the vowels [i, e, a]). Finally, because Wilson's experimental task was a

production task, I included the faithfulness constraint *ALTERNATE, which penalizes any

segment that is unfaithful, in order to capture the fact that participants are generally more

conservative in production tasks (this was motivated in section 4.5.4). Note that my model

therefore had a total of 9 constraints compared to Wilson's 14 constraints.

To get the prior weights for the *MAP constraints, I used the confusion data from Guion

(1998), reported also in Wilson (2006); these are the same confusion data used by Wilson.

Guion's data set reports confusions for [k], [tʃ], [g], and [dʒ] before [i] and [a]. Following

Wilson, I interpolated to get values for the pre-[e] context; in particular, I took the average

number of confusions for the relevant sounds when they occurred before [i] and [a]. Running

these through a MaxEnt model, as described in section 4.3, I got the weights in Table 29, which

were entered as the prior μ values for the constraints in the learning model. A look at the weights

reveals that they reflect the expected similarity relationships: (a) the sounds are more similar

before [i] and least similar before [a], and (b) [k] and [tʃ] are more similar than [g] and [dʒ], all

else equal.

Table 29. Prior weights (μ) for *MAP constraints based on confusion data in Guion (1998).

Constraint	Prior weight (μ)
*MAP(k, tʃ)/_i	0.21
*MAP(k, tʃ)/_e	0.98
*MAP(k, tʃ)/_a	1.87
*MAP(g, dʒ)/_i	1.22
*MAP(g, dʒ)/_e	1.66
*MAP(g, dʒ)/_a	2.27

The two markedness constraints were entered into the model with μ equal to 0. All of the

constraints, with the exception of *ALTERNATE, were assigned a σ^2 of 0.6, the same value used

for my experiments above. *ALTERNATE was assigned a μ of 1.0 and a σ^2 of .0001, just as in my

production study.

The training data for the model was based on the training data that Wilson gave his experimental participants.[39] There were four conditions total; the training data for each condition is summarized in Table 30. Like Wilson, I focused on just two output candidates, the fully faithful candidate and the candidate that has undergone palatalization.

Table 30. Summary of training data entered into the model based on Wilson (2006).

Experiment 1		Experiment 2	
High condition	Mid condition	Voiceless condition	Voiced condition
4 ki → tʃi	4 ke → tʃe	6 ki → tʃi	6 gi → dʒi
6 gi → dʒi	6 ge → dʒe	6 ke → tʃe	6 ge → dʒe
3 ka → ka	3 ka → ka	1 gi → dʒi	1 ki → tʃi
3 ga → ga	3 ga → ga	1 ge → dʒe	1 ke → tʃe
		3 ka → ka	3 ka → ka
		3 ga → ga	3 ga → ga

In Wilson's Experiment 1, participants received training that [k] and [g] got palatalized before [i] (High condition) or before [e] (Mid condition). They also received explicit evidence that [k] and [g] did not palatalize before the low vowel [a]. In Table 31, I report Wilson's results along with the predictions of my model. In the High condition, Wilson found that participants rarely generalized the palatalization to [e], a result that my model is successful in predicting.

In the Mid condition, Wilson predicted that participants would generalize the palatalization to the high vowel context because the velars and palato-alveolars are more similar before [i] than before [e]. Typologically, it is common to palatalize before [i] and [e] or asymmetrically before [i], but not asymmetrically before [e]. In Wilson's actual results, the velars are indeed palatalized

---

[39] Wilson gave his participants a few practice trials. He argues that the practice trials are more salient to participants and, so he weights them heavier in his model. Following suit, I include them in the model's training input and count them as double. For instance, in the High condition of Exp. 1, participants received one case of gi → dʒi in practice, so I included two extra cases (i.e., double) of gi → dʒi in the training input for that condition.

at roughly equal rates before mid vowels and high vowels, as predicted (19% vs. 20% for [k]; 49% vs. 48% for [g]), but there are two odd features of the results for the Mid condition. First, the rate of palatalization for [k] is quite low overall, considering that [ke] → [tʃe] was trained, [k] and [tʃ] are more similar than [g] and [dʒ], and [k] is more likely to palatalize than [g] typologically.[40] Second, and more troubling, is the fact that participants mysteriously generalized to the low vowel context even though they were explicitly trained that palatalization does not occur in that context. In doing so, the participants were going against perceptual similarity, typology, and their training. These problems, which have been pointed out by others (e.g., Moreton & Pater, 2012b), raise serious concerns about the interpretation of the results from that condition.

With those issues in mind, my model does not match Wilson's results closely for the Mid condition of Experiment 1; however, it matches Wilson's *hypothesized* results extremely well: it predicts generalization from the [e] context to the [i] context, but very little generalization to the [a] context.

---

[40] Wilson partly accounts for the [k]/[g] difference by saying that the practice items included a case of [g] → [dʒ] but not a case of [k] → [tʃ], under the assumption that practice items are weighted heavily by participants (see fn. 39 above). This explanation accounts for the high rate of palatalization for [g], but does not really explain the low rate of palatalization for [k].

Table 31. Results from Experiment 1 of Wilson (2006) compared to my model's predictions, in percentage of trials palatalized (experimental results) or predicted to be palatalized (model predictions).

	High condition			Mid condition	
	Model prediction	Experimental result		Model prediction	Experimental result
ki → tʃi	41	44	ki → tʃi	37	20
ke → tʃe	21	13	ke → tʃe	37	19
ka → tʃa	9	5	ka → tʃa	9	15
gi → dʒi	44	52	gi → dʒi	23	48
ge → dʒe	15	14	ge → dʒe	38	49
ga → dʒa	8	14	ga → dʒa	9	39

In Experiment 2, the participants were trained to palatalize either [k] (Voiceless condition) or [g] (Voiced condition) before [i] and [e]. They also received two cases of palatalization of the velar in the opposite voicing category (i.e., [g] in the Voiceless condition, [k] in the Voiced condition), one before [i] and one before [e]. This was to encourage generalization across voicing categories. In both conditions, they were trained that [k] and [g] did not palatalize before [a].

The results are shown in Table 32 along with the predictions of my model. In both conditions, rates of palatalization were moderately high in the cases that were trained (i.e., [ki] → [tʃi] and [ke] → [tʃe] in the Voiceless condition; [gi] → [dʒi] and [ge] → [dʒe] in the Voiced condition). Generalization to the other voicing category was only modest and did not differ substantially across the two conditions. The model predictions are successful at predicting both of these findings in Experiment 2.

Table 32. Results from Experiment 2 of Wilson (2006) compared to my model's predictions, in percentage of trials palatalized (experimental results) or predicted to be palatalized (model predictions).

	Voiceless condition			Voiced condition	
	Model prediction	Experimental result		Model prediction	Experimental result
ki → tʃi	60	39	ki → tʃi	34	26
ke → tʃe	55	36	ke → tʃe	20	20
ka → tʃa	16	12	ka → tʃa	7	0
gi → dʒi	18	14	gi → dʒi	55	50
ge → dʒe	13	11	ge → dʒe	48	44
ga → dʒa	5	9	ga → dʒa	14	23

The overall proportion of variance that was accounted for by the model (r^2) is reported in Table 33 for each of the four conditions. The r^2 reported by Wilson for his substantively biased model is also provided for comparison (i.e., the r^2 for critical test items, which are also the ones that I modeled). Overall, the predictions of my model represent an excellent fit to Wilson's experimental results; the model actually outperforms his model in three of the four conditions. The exception is in the Mid condition of Experiment 1, where my model has essentially no correlation with the experimental results.[41] This condition, however, was the condition in which Wilson's experimental results were problematic, failing to meet the hypothesized pattern, as mentioned above. My model was very successful at predicting the *hypothesized* results. It remains unclear why Wilson's results in that condition were aberrant, and thus whether the model *should* be matching those particular results at all.

---

[41] Despite the r^2 of (virtually) 0, the model's predictions are not crazy. This can be seen in Table 31. When fitting the model to only six observations, the r^2 will be severely affected if a couple of the points are far off.

Table 33. Proportion of variance accounted for (r^2) by Wilson's (2006) substantively biased model and by my substantively biased model, when the model predictions are fitted to Wilson's experimental results (critical test items).

Condition	r^2 reported for Wilson's model	r^2 for my model
Exp. 1 – High condition	.76	.92
Exp. 1 – Mid condition	.58	.002
Exp. 2 – Voiceless condition	.48	.99
Exp. 2 – Voiced condition	.69	.89

In sum, I conclude that my method of implementing the substantive bias, though different from Wilson's, is successful overall in capturing the observations about velar palatalization presented in Wilson (2006).

4.6.3 Initial state

The initial state refers to the (presumably innate) state of the child's grammar before any learning occurs. In a MaxEnt model, the prior is often taken to represent the initial state (e.g., see Goldwater & Johnson, 2003), so it is worth considering what the model presented here assumes about the initial state of the grammar. Of course, the prior in a MaxEnt model is not merely an "initial" state. It represents a bias that continues to affect learning throughout the lifetime, as opposed to a default setting that has no lasting effect once learning has commenced (cf. the GLA; Boersma & Hayes, 2001).

Previous researchers have argued that in the initial state, it must be the case that markedness outranks faithfulness (Gnanadesikan, 1995; Smolensky, 1996; Prince & Tesar, 1999; Boersma & Levelt, 2000; Curtin & Zuraw, 2002; Hayes, 2004; but see Hale & Reiss, 1997, for an opposing point of view). A major argument for having Markedness >> Faithfulness in the initial grammar is that children's early non-adultlike productions appear to reflect principles of markedness, which would be difficult to explain if faithfulness were highly ranked in the grammar.

In the substantively biased instantiation of the model reported above, markedness constraints were set with a μ of 0 whereas the *MAP faithfulness constraints were all set with non-zero prior weights. This choice does appear to bear some importance for the model's performance. Indeed, raising the μ of the markedness constraints to be higher than the highest *MAP μ causes problems for the model's performance – in particular, the weights of the markedness constraints never really have a reason to decrease so the model overgeneralizes.

There is, however, a way to resolve this problem. All of the studies tested here involved alternations in a paradigm (singular/plural forms of nouns), meaning that paradigm uniformity (Steriade, 2000) is in play. Therefore, the *MAP constraints would have the same effect if evaluated as output-output constraints (*MAP-OO) rather than input-output constraints (*MAP-IO). Indeed, considering the constraints to be OO-faithfulness constraints makes more sense in this model because they take perceptual similarity into account; it is odd conceptually to consider the perceptual similarity between an abstract underlying form and a surface form.

Several people have claimed that OO-faithfulness constraints are highly ranked in the initial state because there appears to be a natural tendency to prefer consistent paradigms. Hayes (1997, p. 46), who refers to them as Paradigm Uniformity constraints, sums the argument up as follows:

> I assume that the default position of a Paradigm Uniformity constraint in the grammar is undominated; the language learner expects, a priori, that morphemes will not alternate. The evidence for this is the massive body of evidence that grammatical change tends to be in the direction of "cyclicity" and paradigm leveling."

McCarthy (1998) has also argued that OO-faith ranks high in the initial state, and Tessier (2006, 2012) has provided experimental evidence that children rank OO-faith high in their grammars.

Thus, under the assumption that OO-faith is ranked high in the initial state, the story of acquisition with regards to this model might be as follows. Children are born and immediately begin hearing speech sounds. After months of experience hearing many speech sounds in many environments, they begin to fill in their own P-map with knowledge about the relative similarity of pairs of speech sounds (some may hold that this knowledge is innate, but I think that assumption is not necessary). After learning several words during the first year of life, infants begin learning morphology and start learning that the same lexical items can appear in multiple morphophonological contexts – that is, they start learning paradigms. At this point, they already have a natural preference for paradigm uniformity (i.e., *MAP-OO constraints will be preferentially ranked high as they are induced). These *MAP-OO constraints will be weighted by default based on the P-map that has developed from perceptual experience. The weights can then be altered through learning like any other constraint.

CHAPTER 5

Biased learning of phonological alternations by 12-month-old infants

5.1 Introduction

Children must learn the phonological rules governing phonological alternations like [pæt] ~ [pæɾ] in forms like *pat* and *patting* at some point during the language acquisition process, but few studies have directly investigated how infants acquire such knowledge.

A common view is that infants learn these rules by tracking the distributions of speech sounds and their phonological contexts in their language input. By looking for complementary distributions (i.e., cases where two speech sounds never occur in the same phonological context), infants could discover the cases of systematic variation and acquire the rules that derive them (Peperkamp & Dupoux, 2002; Peperkamp et al., 2006a).

The ability to track distributions based on language input is undeniably a powerful tool available to the language learner. During the first year of life, infants have been shown to use distributional learning for discriminating speech sounds (Maye, Werker, & Gerken, 2002; Anderson, Morgan, & White, 2003; Yoshida et al., 2010), learning phonotactics (Chambers, Onishi, & Fisher, 2003), and segmenting words from running speech (Saffran, Aslin, & Newport, 1996). Indeed, White et al. (2008) found that after brief exposure to an artificial language, 12-month-old infants exposed to [p] only after consonants and [b] only after vowels (e.g., *rot pevi, na bevi...*) treated *poli* and *boli* as variants of the same word, indicating that they had learned the [p ~ b] alternation from distributional evidence.

However, it is unlikely that the unconstrained tracking of distributions is sufficient for properly learning the phonological mappings of a language. For instance, [h] and [ŋ] (the final sound in *sing*) happen to have perfectly complementary distributions in English, but no phonological analysis would claim that one is derived from the other by rule, in part because the two sounds are so phonetically dissimilar.

The experimental results presented in Chapter 3 suggest that adults are indeed biased during learning to avoid saltatory alternations – that is, alternations that "leap over" an intermediate, non-alternating sound. Specifically, adults learning alternations between dissimilar sounds in an artificial language (e.g., [p] changes to [v] between vowels) assumed that intermediate sounds also alternate (e.g., they also changed [b] to [v] between vowels), *without* evidence in the input. They even had a tendency to change intermediate sounds (incorrectly) when presented with direct evidence in the input that they did not change. These observations can be stated another way: participants learning alternations between dissimilar sounds assumed that more similar sounds also alternated, consistent with principle of minimal modification in Steriade's theory of the P-map (2001/2008).

However, it remains unclear whether the bias exhibited by adults is active in early language acquisition. Adult language learners are able to draw on experience and problem-solving strategies that are not available to infants. In the present study, we investigate whether 12-month-olds exhibit such a bias.

5.2 Experiment

Our testing paradigm was based on White et al.'s (2008) study showing that 12-month-olds (but not 8.5-month-olds) can learn novel alternations after brief exposure to an artificial

language. In the DISSIMILAR-TO-SIMILAR condition, we exposed 12-month-old infants to alternations involving dissimilar sounds: voiceless stops and voiced fricatives (i.e., [p ~ v] or [t ~ z]). Infants were exposed to words (e.g., *poli*) preceded by the short "function" words *na* or *rom* (e.g., *rom poli*). For each infant, sounds at one place of articulation (labials or coronals) were in complementary distribution; voiced fricatives (e.g., [v]) only appeared after *na* and voiceless stops (e.g., [p]) only appeared after *rom*—providing distributional evidence for a phonological alternation. Sounds at the other place of articulation (e.g., [t] and [z]) were non-alternating (i.e., contrastive), appearing after both *na* and *rom*.

Based on White et al.'s (2008) results, infants hearing [p] and [v] in complementary distribution and [t] and [z] with overlapping distributions should assume that, e.g., *puni* and *vuni* are context-dependent variants of the same word, but *tilu* and *zilu* are different words. The goal of the current experiment was to determine whether infants exposed to alternations between dissimilar sounds assume, like adults, that more similar sounds also alternate. Thus, after training infants that [p] and [v] alternate and [t] and [z] do not alternate (or vice-versa), we tested them on pairs containing the intermediate sounds [b] and [d] - *sounds they had never heard during the exposure phase*. Note that the general concept of this experiment is similar to that of Experiment 1 with the adults (section 3.4): train on potentially saltatory alternations and then test on withheld intermediate sounds.

If infants have a bias like the adult learners, when they learn that dissimilar sounds like [p] and [v] alternate, they should assume that [b] and [v] also alternate; but they would have no reason to think that [d] and [z] alternate. Thus, at test they should treat pairs beginning with labials (e.g., *buni/vuni*) differently than those beginning with coronals (e.g., *dilu/zilu*). To ensure that infants only generalize from dissimilar sounds to more similar sounds, we ran a control

SIMILAR-TO-DISSIMILAR condition where infants were exposed to alternations between similar sounds ([b ~ v], [d ~ z]) and then tested on dissimilar sounds ([p] and [t]). Because infants had never heard [p] or [t] during exposure, and they were not intermediate between the learned alternations, infants had no reason to treat pairs of words beginning in labials and coronals any differently in the SIMILAR-TO-DISSIMILAR condition.

5.2.1 Method

5.2.1.1 Participants

Forty 12-month-old infants (26 males, 14 females, mean age = 370 days, age range = 349 – 407 days) participated in the study. All infants had more than 85% input in English based on a parental language questionnaire (Bosch & Sebastián-Gallés, 2001; Sundara & Scutellaro, 2011). Eleven additional infants were tested but excluded due to crying (n=9), experimenter error (n=1), or equipment problems (n=1).

5.2.1.2 Design and stimuli

The experiment consisted of two phases: an exposure phase and a test phase. During the exposure phase, infants heard several phrases in an artificial language (e.g., *na poli*), each consisting of a monosyllabic word (*na* or *rom*) followed by one of sixteen disyllabic words. Infants were randomly assigned to either the DISSIMILAR-TO-SIMILAR condition or the SIMILAR-TO-DISSIMILAR condition depending on whether the alternation they learned involved dissimilar sounds or similar sounds (sample stimuli in Table 34).

In the DISSIMILAR-TO-SIMILAR condition, the disyllabic words always began with either [p], [v], [t], or [z]. Infants were randomly assigned to one of two sub-groups. For each sub-group,

consonants at one place of articulation (labials or coronals) were in complementary distribution and consonants at the other place of articulation had overlapping distributions. For example, the Labials sub-group heard [p]-initial words only after *rom* and [v]-initial words only after *na,* but they heard [t]- and [z]-initial words after both *na* and *rom.* Based on this distribution, word-initial labials alternated depending on context—[v] only appeared after vowels, [p] only after consonants—but word-initial coronals were contrastive because they appeared in both contexts (and vice-versa for the Coronals sub-group).

In the test phase, infants heard novel pairs of words repeated <u>without</u> *na* or *rom.* As in White et al., (2008), we did not include the conditioning context for the alternation (i.e., the "function" words) at test, to be sure that infants were successful at associating the variant forms with a single base form, rather than merely relying on transitional probabilities. Infants in both the Labials and Coronals sub-groups heard the same set of test trials, half of which contained words beginning with labials and half beginning with coronals. In this phase, words began with voiced stops and voiced fricatives. Crucially, voiced stops were sounds that infants had <u>never</u> heard during exposure. Note that in the DISSIMILAR-TO-SIMILAR condition, the two voiced stops [b, d] were intermediate between the initial sounds in the training words ([b] is intermediate between [p] and [v], [d] is intermediate between [t] and [z]). For each group of infants, one voiced stop was intermediate between the alternating sounds (Alternating trials) and the other voiced stop was intermediate between contrastive sounds (Contrastive trials). For each infant, which test trials were Alternating trials and which trials were Contrastive trials depended on her sub-group (Labials or Coronals) during the exposure phase.

For the control SIMILAR-TO-DISSIMILAR condition, the same set of words was used, except the exposure words began with [b] and [d] (rather than [p] and [t]) and the test words began with

[p] and [t] (rather than [b] and [d]). In this condition, the novel sounds at test ([p] and [t]) were not intermediate between the word-initial sounds during training. A full list of stimuli is provided in the Appendix (section 5.4).

Table 34. Example stimuli to illustrate the experimental design. Note that infants heard more phrases during the exposure phase than are listed here (see Appendix). Shaded cells mark alternating forms.

		DISSIMILAR-TO-SIMILAR condition		SIMILAR-TO-DISSIMILAR condition	
	Labials sub-group		Coronals sub-group	Labials sub-group		Coronals sub-group
Exposure phrases	*rom poli* *na voli* *rom poli* *na voli*	*rom poli* *rom voli* *na poli* *na voli*	*rom boli* *na voli* *rom boli* *na voli*	*rom boli* *rom voli* *na boli* *na voli*
	rom timu *rom zimu* *na timu* *na zimu*	*rom timu* *na zimu* *rom timu* *na zimu*	*rom dimu* *rom zimu* *na dimu* *na zimu*	*rom dimu* *na zimu* *rom dimu* *na zimu*
Test pairs	*buni/vuni, bagu/vagu, dilu/zilu, dari/zari*		*puni/vuni, pagu/vagu, tilu/zilu, tari/zari*	

A female native English speaker, who had phonetic training but was unfamiliar with the purpose of the study, produced the stimuli. The recording was done using PcQuirerX (sampling rate 44,100 Hz) in a soundproof booth using a Shure SM10A head-mounted microphone, whose signal ran through an XAudioBox pre-amplifier and A-D device box. The stimuli were recorded naturally, as two word phrases for the exposure stimuli and as single words for the test stimuli, using an infant-directed speaking style. Stress was placed on the first syllable of the disyllabic word. The initial stress of the disyllabic word, as well as the low transitional probabilities between the monosyllabic "function" word and the disyllabic word, were intended to help infants posit a word boundary between the monosyllabic word and the disyllabic word (e.g., Saffran, Aslin, & Newport, 1996; Thiessen & Saffran, 2003).

5.2.1.3 Apparatus

Infants were seated on their caregiver's lap approximately 3.5 feet from a display monitor in a curtained soundproof booth. The auditory stimuli were played at a comfortable 78 dB volume over JBL speakers located just next to the monitor. Presentation of stimuli and data recording were handled automatically by Habit X (Cohen, Atkinson, & Chaput, 2004).

The experimenter sat in an adjacent room watching the infant via a monitor connected to a Sony digital video camera hidden just under the display screen in front of the infant. Both the experimenter and the caregiver wore headphones playing music so they could not influence the infant's behavior.

5.2.1.4 Procedure

Infants were tested using the visual fixation procedure (Werker et al., 1998). At the beginning of each trial, a looming light was paired with a baby giggle to attract the infant's attention. When the infant looked at the screen, a picture of a flower appeared on the screen while an auditory stimulus was played simultaneously over the speakers. The same flower appeared for all exposure trials and a different flower appeared for all test trials.

In the exposure phase, infants heard three trials lasting 45 seconds each for a total of 135 seconds. Each exposure trial contained all of the exposure phrases (e.g., *na voli, rom tago, rom poli...*), with a 300 ms pause between each phrase. Each disyllabic "content" word was presented twice per trial, either once with each monosyllabic word for contrastive words, or twice with the same monosyllabic word for alternating words. The order of the phrases was randomized in each of the three trials. The exposure trials were not contingent on infant looking to ensure that each

infant had the same amount of exposure. The three trials were presented in a random order by the experimental software.

In the test phase, infants in both the Labials and Coronals sub-groups heard the <u>same</u> test trials (3 blocks X 4 trials = 12). The test trials were fully contingent on infant looking. The next test trial began either after the infant had looked away from the screen for more than one second or after the maximum test trial duration (20 seconds). A trial was repeated if the infant looked away during the first two seconds of the trial.

Each trial contained one pair of test words repeated several times (e.g., *bagu, vagu, vagu, bagu...*) with a 300 ms pause between each word. Within a trial, order of the words was pseudo-randomized, with the restriction that words could occur only twice in a row and both words appeared as one of the first two words of each trial. Each pair of test words was presented once per block. Order of the test trials was counterbalanced across infants.

5.2.2 Results

No significant effects were found based on sub-group (Labials vs. Coronals), so they have been collapsed in the analysis. The results were analyzed using a 2 x 2 mixed ANOVA with a between-subjects variable for Condition (DISSIMILAR-TO-SIMILAR or SIMILAR-TO-DISSIMILAR), a within-subjects variable for Trial Type (Alternating vs. Contrastive), and looking time (in seconds) as the dependent variable (Figure 15). The ANOVA revealed no main effect of Condition, $F(1, 38) = 1.390$, $p = .25$, $\eta^2_p = .04$, and no main effect of Trial Type, $F(1, 38) = .007$, $p = .93$, $\eta^2_p = 0$, but there was a significant Condition by Trial Type interaction, $F(1, 38) = 5.315$, $p = .027$, $\eta^2_p = .12$.

Post-hoc paired-samples t-tests (with Bonferroni-adjusted alpha levels of .025) showed that in the DISSIMILAR-TO-SIMILAR condition, infants looked significantly longer to Contrastive trials ($M = 8.34$, $SD = 3.36$) than to Alternating trials ($M = 7.35$, $SD = 3.02$), $t(19) = 3.364$, $p < .01$, Cohen's $d = .31$. However, in the SIMILAR-TO-DISSIMILAR condition, there was no difference in looking time between the Contrastive trials ($M = 7.60$, $SD = 3.34$) and the Alternating trials ($M = 7.92$, $SD = 3.44$), $t(19) = .658$, $p = .52$, Cohen's $d = .09$.

Figure 15. Mean looking time (in sec) for the Alternating trials and the Contrastive trials, according to Condition. Error bars show standard error of the mean.
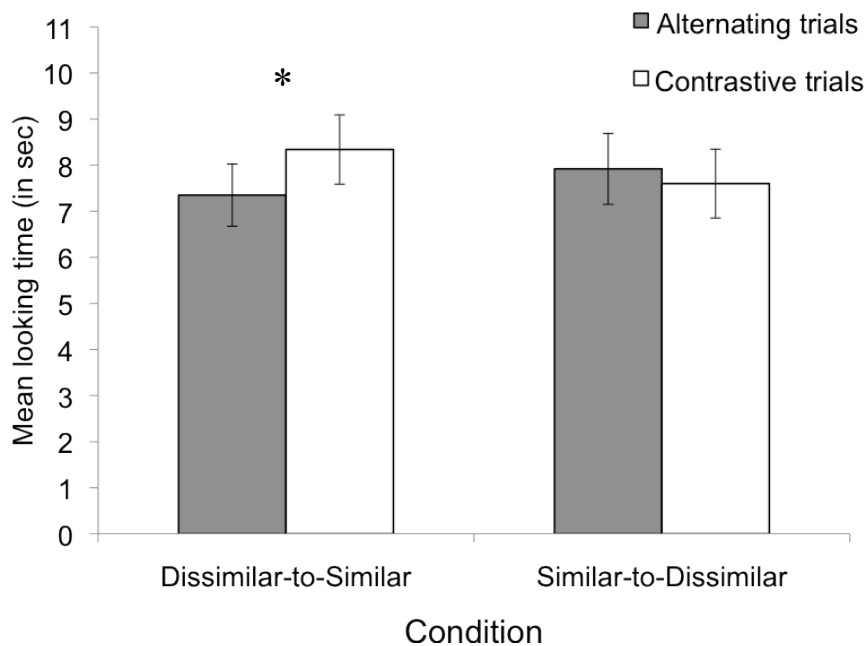
5.3 Discussion

As predicted, 12-month-olds who learned alternations between dissimilar sounds (DISSIMILAR-TO-SIMILAR condition), but not between similar sounds (SIMILAR-TO-DISSIMILAR condition) differentiated between Alternating and Contrastive trials. Recall that infants had no

evidence from the input that could have led them to treat the Alternating and Contrastive trials differently - all test trials had novel sounds not presented during the exposure phase. Despite the lack of evidence in the input, infants in the DISSIMILAR-TO-SIMILAR condition treated novel sounds that were <u>intermediate between the alternating sounds</u> (Alternating trials) differently from those that were <u>intermediate between contrastive sounds</u> (Contrastive trials). In the Alternating trials, potential saltation was at stake, but not in the Contrastive trials. These results have several implications for understanding how phonological alternations are acquired.

First, these results provide corroborating evidence that 12-month-old infants can learn novel phonological alternations after brief exposure to an artificial language (as found by White et al., 2008). Only if infants had learned the alternations presented during exposure would we expect differences between test items in either condition. It is worth noting that the direction of the effect was different in the current study and in White et al.'s study: we found that infants listened longer on non-alternating trials whereas White et al. found that they listened longer on alternating trials. This difference is plausibly due to the fact that White et al.'s study included no novel sounds at test whereas the current study tested infants on novel word-initial sounds.

Second, the results show that infants tacitly make assumptions when learning alternations. We argue that the infants who learned *poli* alternates with *voli*, assumed that because [b] is intermediate between [p] and [v], it is likely that *buni* alternates with *vuni* as well. Due to this indirect evidence, the *buni/vuni* trials were not so novel after all, but the *dilu/zilu* trials were fully novel. Stated another way, infants learning alternations between dissimilar sounds (e.g., [p ~ v]) generalized the alternations to sounds that were more similar (e.g., [b ~ v]). This generalization was only in one direction: from less similar sounds to more similar sounds. Infants learning alternations between similar sounds did not generalize to dissimilar sounds (recall that there was

177

no difference between test trials in the SIMILAR-TO-DISSIMILAR condition). Thus, like adult learners, infants exhibit a bias when learning alternations.

These results add to a growing body of literature showing that phonological learning is constrained by certain biases (e.g., Peperkamp, Skoruppa, & Dupoux, 2006b; Wilson, 2006; Finley & Badecker, 2008; Moreton, 2008; Hayes et al., 2009; Becker, Ketrez, & Nevins, 2011; Hayes & White, 2013). Few previous studies have investigated specific biases in infant phonological learning. Jusczyk, Smolensky, and Allocco (2002) argued that infants have an innate bias to prefer "unmarked" forms (in particular, forms with assimilated nasals). On the other hand, Seidl and Buckley (2005), who note that Jusczyk et al.'s methodology was flawed, found that 9-month-olds showed no difference in learning phonetically motivated patterns and arbitrary patterns. In light of these conflicting results, the current study represents a significant advancement by demonstrating a specific case where infants exhibit a learning bias when acquiring novel phonological alternations.

What do the results mean for the language learner? The results suggest that, like adults, infant learners have a soft bias: they avoid saltations, and more generally, they assume that alternations between similar sounds are more likely than alternations between dissimilar sounds. The results are consistent with Steriade's (2001/2008) claim that learners assume phonological processes will involve minimal modification, as implemented in the learning model in Chapter 4.

5.4 Appendix

Full list of stimuli.

DISSIMILAR-TO-SIMILAR condition:

	Labials sub-group				Coronals sub-group			
rom	*poli*	*panu*	*pezi*	*pika*	*tovi*	*tago*	*turo*	*timu*
na	*voli*	*vanu*	*vezi*	*vika*	*zovi*	*zago*	*zuro*	*zimu*
rom/na	*tovi*	*tago*	*turo*	*timu*	*poli*	*panu*	*pezi*	*pika*
	zovi	*zago*	*zuro*	*zimu*	*voli*	*vanu*	*vezi*	*vika*
Test pairs	*buni/vuni, bagu/vagu, dilu/zilu, dari/zari*							

SIMILAR-TO-DISSIMILAR condition:

	Labials sub-group				Coronals sub-group			
rom	*boli*	*banu*	*bezi*	*bika*	*dovi*	*dago*	*duro*	*dimu*
na	*voli*	*vanu*	*vezi*	*vika*	*zovi*	*zago*	*zuro*	*zimu*
rom/na	*dovi*	*dago*	*duro*	*dimu*	*boli*	*banu*	*bezi*	*bika*
	zovi	*zago*	*zuro*	*zimu*	*voli*	*vanu*	*vezi*	*vika*
Test pairs	*puni/vuni, pagu/vagu, tilu/zilu, tari/zari*							

179

CHAPTER 6

General conclusions

6.1 Summary of the dissertation

The goals of this dissertation were to investigate phonological learning by understanding a single difficult issue, saltation, from multiple angles. Over the course of the dissertation, I have explored saltation from several perspectives; the main findings are reviewed below. Overall, it is clear that the findings in this dissertation have implications for phonological theory and phonological learning. By undertaking a broad investigation of this particular phenomenon, we have gained insights about the nature of the phonological grammar and the role that perceptual similarity may play in constraining phonological learning.

In Chapter 2, we saw that saltation is attested in several languages, but that overall it appears to be cross-linguistically rare. Of the cases that are attested, most of them apply only to single segments rather than classes of sounds, and many have other peculiarities that bring into question their status as productive phonological processes of the language. I conjectured (following Minkova, 1991 and Lass, 1997) that saltations are never innovated directly through sound change, but arise instead by accident from a series of independent, but non-saltatory, sound changes. I also presented a brief overview of historical evidence supporting this view for several of the attested cases. Finally, I showed that saltation is not derivable in classical OT (Prince & Smolensky, 1993/2004) and other closely related frameworks, primarily because of the set of feature-based faithfulness constraints assumed in traditional Correspondence Theory (McCarthy & Prince, 1995). I argued that a solution to saltation involving local constraint conjunction, as

proposed by Lubowicz (2002) and Ito and Mester (2003), was not ideal because of the theoretical baggage associated with adding conjoined constraints to the theory. Instead, I proposed an analysis involving *MAP faithfulness constraints (Zuraw, 2007), constrained by a substantive bias based on the P-map (Steriade, 2001/2008).

In Chapter 3, we saw from two artificial language experiments with adults that saltation is a dispreferred pattern for learners. Participants trained on data that were ambiguous between a saltatory system and a non-saltatory system preferred the non-saltatory interpretation (Experiment 1). Even when they were trained on explicit cases of saltation, participants found the saltation difficult to learn compared to comparable non-saltatory cases (Experiment 2).

In Chapter 4, I began by presenting the challenge to phonological theory posed by saltation: it is dispreferred, yet ultimately learnable. To account for these observations, I proposed a phonological framework with three components: *MAP faithfulness constraints penalizing correspondences between individual segments (from Zuraw, 2007), a substantive bias based on the P-map (from Steriade, 2001), and a MaxEnt learning framework for implementing the bias (e.g., Goldwater & Johnson, 2003; Wilson, 2006). We saw that the predictions of the substantively biased model provided an excellent fit to experimental results, crucially accounting for the anti-saltation effect observed in the experiments. The substantively biased model outperformed an unbiased model as well as a model in which the *MAP constraints had high prior weights but did not differ according to perceptual similarity, suggesting that a substantive bias based on the P-map plays a role in the learning of phonological alternations. According to the model, learners assign higher *a priori* likelihoods to alternations between similar sounds than to alternations between dissimilar sounds.

In Chapter 5, we saw results from an artificial language experiment with 12-month-olds suggesting that infants learning novel alternations between dissimilar sounds generalized the alternations to affect more similar sounds. However, the reverse was not true; infants learning alternations between similar sounds did not generalize to dissimilar sounds. These results provide evidence that the minimal modification bias is also active during early language acquisition, not only in experiments with adult learners.

Overall, the results from this dissertation tell a coherent story. Saltation is a relatively uncommon phenomenon in the world's languages. Learners disprefer saltatory phonological systems due to a bias based on the theory of the P-map, which causes them to prefer alternations with minimal modification. When saltatory systems arise in languages, they do so by historical accident and appear to be unstable over several generations, at least in some cases.

I conclude by considering potential implications of this research in the areas of phonological theory, phonological acquisition, and language change.

6.2 Implications for phonological theory

The findings in this dissertation have important implications for the type of framework that is needed in phonological theory. First, it seems clear that we must extend the constraint inventory beyond the traditional set of constraints assumed in classical OT (Prince & Smolensky, 1993/2004; McCarthy & Prince, 1995), which are insufficient for deriving saltation. Conjoining markedness and faithfulness constraints is one solution to the problem, but the cost to phonological theory may be too high (see section 2.5.1; also Ito & Mester, 1998). Instead, a *MAP family of constraints (Zuraw, 2007), which are constrained in a principled way based on

the theory of the P-map (Steriade, 2001/2008), strikes what seems to be the correct balance between power and restrictiveness.

Second, this work has provided new evidence for the role of substantive bias, particularly on the basis of perceptual similarity, in phonological learning. As we saw, an account based on featural complexity fails to account for all of the nuances in the experimental results, such as the preference for [b ~ v] over [f ~ v] with no training on [b] or [f]. The substantively biased model, by contrast, provides an excellent fit to the experimental results, outperforming comparable models without a substantive bias. In addition, this research underscores the usefulness of MaxEnt grammar models for investigating questions related to biased phonological learning, as Wilson (2006) showed previously.

This phonological framework also suggests new potential analyses of certain phonological phenomena. For instance, consider the case of synchronic chain shifts (e.g., /p/ → [b], /b/ → [v]), also known as "counterfeeding on the focus" (McCarthy, 1999). Such patterns pose problems for classical OT because the relevant generalizations are not true on the surface: if a highly ranked markedness constraint motivates hypothetical /aba/ → [ava], why does /apa/ surface as [aba] rather than [ava]? However, such cases are not problematic for OT as long as some constraint can rule out the "one fell swoop" candidate, i.e. /apa/ → [ava] (McCarthy, 1999; Baković, 2007).

One analysis of such cases relies on the local conjunction of two faithfulness constraints (Kirchner, 1996). For example, the conjoined constraint [ID(voice) & ID(cont)] could serve to keep /p/ from changing all the way to [v]. However, under the *MAP + P-MAP framework proposed here, an analysis of chain shifts does not require conjoined constraints because highly weighted *MAP constraints (e.g., *MAP(p, v)) can serve to rule out the "one fell swoop"

candidate. In spirit, this analysis has similarities to Gnanadesikan's (1997) analysis of chain shifts, which involved faithfulness scales.

In addition, the MaxEnt model can be used to generate specific predictions about the learnability of these patterns. For instance, synchronic chain shifts should be easy to learn because they are P-map compliant, and thus consistent with the model's prior: large changes are not allowed (due to *MAP constraints with substantial weights) but smaller changes are acceptable (due to *MAP constraints with smaller weights). Learning outcomes that do not require working *against* the prior should be learned with relative ease.

Finally, the *MAP + P-MAP framework has potential implications for how we think of the relative contributions of markedness and faithfulness in phonology. Specifically, under this theory, some of the typological work that used to be done by markedness constraints is (potentially) shifted to faithfulness and substantive bias. For instance, consider the case of velar palatalization considered by Wilson (2006) and reanalyzed here (see section 4.6.2). Traditionally, the cross-linguistic tendency to find palatalization before high vowels but not before low vowels was considered an issue of markedness: there is some markedness constraint, e.g. *ki, that is preferentially ranked higher than another markedness constraint, *ka (or perhaps *ka does not exist at all in the universal set of constraints). In other words, the sequence [ki] may be considered more marked than the sequence [ka].

By contrast, in the *MAP + P-MAP approach, the tendency to have palatalization before high vowels but not before low vowels is attributed to faithfulness and the P-map. [k] and [tʃ] are more perceptually similar before [i] than before [a], so there is a greater bias to be faithful (i.e., to avoid mappings between [k] and [tʃ]) before [a]. As a result, the analysis will work even if markedness constraints are much simpler. There is no need for *ki, *k[–low], etc. to derive the

184

pattern; all that is needed to motivate the change is *kV, or perhaps even *[+velar]V. Note that even though Wilson also implemented a MaxEnt model with a substantive bias, his model does not have this property; all of the work is done in the markedness constraints. It is an open question whether we *want* our theory of phonology to behave in this way. I leave this question, as well as a deeper investigation of this property of the theory, for future inquiry.

6.3 Implications for phonological acquisition

These findings also have implications for the time course of early language acquisition, especially in terms of learning phonological alternations. First, the results of the infant experiment (Chapter 5) provide corroborating evidence that 12-month-olds can learn novel phonological alternations after brief exposure to an artificial language, as was originally found by White et al. (2008). Only if infants could learn the alternations that they were trained on would we have expected any effects in the experiment. Given that infants can indeed learn alternations with such little exposure (i.e., about 2 minutes) in a new language, this raises the question of whether younger infants have already picked up on at least some of the alternations of their native language. White et al. found that 8.5-month-olds could not learn the alternations in their artificial language study, but we still know very little about when infants learn the alternations of their own language.

Second, the results of the adult artificial language experiments (Chapter 3) and the infant experiment (Chapter 5) indicate that infants should find it difficult to learn saltations. Moreover, consider a language with alternations in which two distinct phonemes neutralize to a single sound, such as tapping in American English where /t/ and /d/ both neutralize to [ɾ]. These findings suggest that an infant who has learned the alternation between the more dissimilar pair

185

of sounds will necessarily already have learned the alternation between the more similar pair of sounds. Concretely, once American English learning infants have learned /t/ → [ɾ], they should already have learned /d/ → [ɾ]. Otherwise, they would have learned a saltation unnecessarily.

More generally, we can make the further prediction that infants will learn alternations between more similar sounds earlier than those between less similar sounds, all else being equal. This follows from the principle of minimal modification expressed in Steriade's P-map theory (Steriade, 2001/2008) as it is implemented via the prior in the learning model proposed here. Alternations between dissimilar sounds have to overcome a larger prior *MAP weight, meaning that they require more supporting data in the input to learn. These predictions suggest several possible directions for future work.

6.4 Implications for language change

Linguists have long been interested in which factors influence the course of language change, and relatedly, in which factors have led to the typological generalizations that we find when looking across the world's languages. We find that certain phonological patterns occur frequently in language after language, other patterns occur in only a handful of languages, and some logically possible patterns are unattested in any known language. These observations have led researchers to propose a number of explanations for the preponderance or rarity of certain phonological patterns.

One may assume that cross-generational language change occurs when children acquire a slightly different grammar than the one internalized by the adults who provided the input (Kiparsky, 1982). The new generation's grammar may be radically different than the one acquired by adults, as in cases of phonological restructuring (e.g., Bowers, 2012), but it is likely

186

more often due to subtle differences in the relative frequencies of the possible output forms generated by the grammar.

There are several possible reasons that multiple languages might, over the course of many generations, tend towards the same phonological patterns to the exclusion of other patterns. Some theories claim that certain patterns are common due to universal phonological knowledge shared by all humans, such as having a universal set of constraints, as assumed in classical OT (Prince & Smolensky, 1993/2004; see also Becker et al., 2011; Becker et al., 2012). Another possibility is that the nature of the human articulatory and/or perceptual system makes it such that certain sound changes are bound to occur repeatedly in language after language whereas other sound changes are highly unlikely to occur naturally. For instance, phrase-final devoicing, which occurs in many languages, might arise naturally because articulatory limitations make it difficult to maintain voicing in word-final obstruents and perceptual limitations may lead listeners to frequently mishear final voiced obstruents as voiceless (e.g., see Blevins, 2004, 2006; Ohala, 1983, 1997).

In addition to these possibilities, several researchers have proposed that the mechanisms involved in language acquisition themselves have a role to play in guiding language change (e.g., Kiparsky, 1982; Clark & Roberts, 1993; Hudson Kam & Newport, 2009; Martin, 2011). Under such accounts, the biases that learners bring to the acquisition process may make some patterns more difficult to learn than others, or else lead learners towards certain outcomes. Over time, dispreferred grammars would be learned less successfully, which would cause the dispreferred patterns to gradually give way, over several generations, to ones that are easier to learn.

The bias implemented here might be one such learning bias. If so, the model presented here makes two predictions about language change. First, it predicts that phonological processes

should tend towards minimal modification because minimal processes are preferred by the learner. This prediction is consistent with the basic idea of P-map as proposed by Steriade (2001/2008).

Second, saltations should be rare in the world's languages, and when they occur, the saltatory system should be somewhat unstable, with a tendency to change to a non-saltatory system over time. Historical evidence (reviewed in section 2.3; see also Hayes & White, in prep.) suggests that this prediction is also on the right track: attested saltations appear to have arisen not directly, but as a result of several independent, non-saltatory sound changes, resulting in a system that was restructured to be saltatory. Further evidence (e.g., Crosswhite, 2000) suggests that these saltatory systems are unstable and tend to lose their productivity over time. Of course, further work should be done before we can say conclusively whether or not these predictions are borne out, particularly when it comes to the long-term (in)stability of saltatory phonological systems.

References

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech, 46*(2–3), 155–182.

Anderson, S. (1981). Why phonology isn't "natural". *Linguistic Inquiry, 12,* 493–547.

Anttila, A. (1997). *Variation in Finish phonology and morphology*. Ph.D. dissertation, Stanford.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R.* Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

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.

Baković, E. (2007). A revised typology of opaque generalisations. *Phonology, 24*(2), 217–259.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-28. http://lmer.4-forge.r-project.org/.

Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic 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.

Benua, L. (1997). *Transderivational Identity: Phonological Relations Between Words*. Ph.D. dissertation, University of Massachusetts, Amherst.

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

Berger, A. L., Della Pietra, V. J., & Della Pietra, S. A. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics, 22*(1), 39–71.

Bevington, G. L. (1974). *Albanian Phonology.* Otto Harrassowitz: Wiesbaden.

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

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics, 32,* 117–165.

Blevins, J., & Garrett, A. (2004). The evolution of metathesis. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 117–156). Cambridge: Cambridge University Press.

Boersma, P. (1997). How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, 21,* 43–58.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives.* The Hague: Holland Academic Graphics.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry, 32*(1), 45–86.

Boersma, P., & Levelt, C. (2000). Gradual constraint-ranking learning algorithm predicts acquisition order. *Proceedings of Child Language Research Forum, 30* (pp. 229–237). Stanford.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.44, retrieved 20 October 2010 from http://www.praat.org.

Bolognesi, R. (1998). *The phonology of Campidanian Sardinian: A unitary account of self-organizing structure.* The Hague: Holland Academics.

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2*(1), 29–49.

Bowers, D. (2012). *Phonological restructuring in Odawa*. M.A. thesis, UCLA.

Bradshaw, M. (1995). Tone on verbs in Suma. In A. Akinlabi (Ed.), *Theoretical Approaches to African Linguistics* (pp. 255–272). Trenton, NJ: Africa World Press.

Bradshaw, M. (1998). Tone alternation in the associative construction of Suma. In I. Maddieson & T. J. Hinnebusch (Eds.), *Language History and Linguistic Description in Africa* (pp. 117–125). Trenton: Africa World Press.

Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language, 58*(2), 265–289.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*(2), 69–77.

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

Clark, R., & Roberts, I. (1993). A computational model of language learnability and language change. *Linguistic Inquiry, 24*(2), 299–345.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0).* Austin: University of Texas.

Côté, M.-H. (2000). *Consonant cluster phonotactics: A perceptual approach.* Ph.D. dissertation, MIT.

Côté, M.-H. (2004). Syntagmatic distinctness in consonant deletion. *Phonology, 21,* 1–41.

Cristia, A., Daland, R., Mielke, J., & Peperkamp, S. (in press). Constrained generalization of implicitly learned sound patterns. *Laboratory Phonology.*

Cristià, A., & Seidl, A. (2008). Phonological features in infants' phonotactic learning: Evidence from artificial grammar learning. *Language, Learning, and Development, 4*(3), 203–227.

Crosswhite, K. M. (2000). Vowel reduction in Russian: a unified account of standard, dialectal, and "dissimilative" patterns. In K. M. Crosswhite & J. McDonough (Eds.), *University of Rochester Working Papers in the Language Sciences, Vol. Spring 2000, No. 1* (pp. 107–172).

Curtin, S., & Zuraw, K. (2002). Explaining constraint demotion in a developing system. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *BUCLD 26: Proceedings of the 26ᵗʰ Annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusion by native and non-native listeners. *Journal of the Acoustical Society of America, 116*(6), 3668–3678.

Eisner, J. (2000). Review of Kager: "Optimality Theory". *Computational Linguistics, 26*(2), 286–290.

Finley, S. (2008). *The formal and cognitive restrictions on vowel harmony*. Ph.D. dissertation, Johns Hopkins.

Finley, S., & Badecker, W. (2008). Substantive biases for vowel harmony languages. In J. Bishop (Ed.), *Proceedings of the West Coast Conference on Formal Lingusitics 27,* 168–176.

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.

Fleischhacker, H. A. (2005). *Similarity in phonology: Evidence from reduplication and loan adaptation*. Ph.D. dissertation, UCLA.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology, 18*, 7–44.

Flemming, E. (2012). Violations are ranked, not constraints: A revised model of constraint interaction in phonology. Handout from a talk given at Rutgers University. March 3, 2012.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61*, 191–218.

Gnanadesikan, A. (1995). Markedness and faithfulness constraints in child phonology. Ms., University of Massachusetts, Amherst. [ROA 67].

Gnanadesikan, A. E. (1997). *Phonology with ternary scales*. Ph.D. dissertation, UMass Amherst.

Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory*.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*(5), 178–186.

Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica, 55,* 18–52.

Hale, M., & Reiss, C. (1997). Grammar optimization: The simultaneous acquisition of constraint ranking and a lexicon. Ms., Concordia University. [ROA 231].

Hayes, B. (1997). Anticorrespondence in Yidiny. Ms., UCLA.

Hayes, B. (1999). Phonetically-driven phonology: the role of optimality theory and inductive grounding. In M. Darnell, E. Moravcsik, M. Noonan, F. Newmeyer, & K. Wheatly (Eds.), *Functionalism and formalism in linguistics, Volume I* (pp.243–285). Amsterdam: John Benjamins.

Hayes, B. (2000). Gradient well-formedness in Optimality Theory. In J. Dekkers, F. van der Leeuw, & J. van de Weijer (Eds.), *Optimality Theory: Phonology, Syntax, and Acquisition* (pp. 88–120). Oxford: Oxford University Press.

Hayes, B. (2009). *Introductory Phonology.* West Sussex: Wiley-Blackwell.

Hayes, B., Kirchner, R., & Steriade, D. (Eds.). (2004). *Phonetically Based Phonology.* Cambridge: Cambridge Univ. Press.

Hayes, B. & MacEachern, M. (1998). Quatrain form in English folk verse. *Language, 64,* 473–507.

Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry, 44,* 45–75.

Hayes, B. & White, J. (in prep.). Saltation and the P-map. Ms., UCLA.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry, 39,* 379–440.

Hayes, B., Wilson, C., & Shisko, A. (2012). Maxent grammars for the metrics of Shakespeare and Milton. *Language, 88,* 691–731.

Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language, 85,* 822–863.

Hellberg, S. (1978). Unnatural phonology. *Journal of Linguistics, 14,* 157–177.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59,* 30–66.

Hutchinson, J. P. (1981). *The Kanuri Language.* Madison, WI: African Studies Program, University of Wisconsin.

Ito, J. & Mester, A. (1997). Correspondence and compositionality: The *ga-gyō* variation in Japanese phonology. In I. Roca (Ed.), *Constraints and Derivations* (pp. 419–462). Oxford: Oxford University Press.

Ito, J. & Mester, A. (1998). Markedness and word structure: OCP effects in Japanese. Ms., UCSC. [ROA 225].

Ito, J., & Mester, A. (2003). On the sources of opacity in OT: Coda processes in German. In C. Féry & R. van de Vijver (Eds.), *The Syllable in Optimality Theory* (pp. 271–303), Cambridge: Cambrige Univ. Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446.

Jarrett, K. (2007). *A Dictionary of Manga, a Kanuri Language of Eastern Niger and NE Nigeria.* Unpublished PDF file, revised and edited by R. Blench. Available at http://www.rogerblench. info/Language/Nilo-Saharan/Saharan/Manga%20dictionary%20Unicode.pdf.

Johnson, M. (2002). Optimality-theoretic lexical functional grammar. In P. Merlo & S. Stevenson (Eds.), *The Lexical Basis of Sentence Processing: Formal, Computational and Experimental Issues* (pp. 59–74). Amsterdam: John Benjamins.

Jun, J. (2004). Place assimilation. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 58–86). Cambridge: Cambridge Univ. Press.

Jusczyk, P. W., Smolensky, P., & Allocco, T. (2002). How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition, 10,* 31–73.

Kiparsky, P. (1973). Abstractness, opacity, and global rules. In O. Fujimura (Ed.), *Three Dimensions of Linguistic Theory* (pp. 1–136). Tokyo, Taikusha.

Kiparsky, P. (1982). *Explanation in Phonology.* Dordrecht: Foris.

Kirchner, R. (1996). Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry, 27*(2), 341–350.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*(3), 245–294.

Lass, R. (1997). *Historical Linguistics and Language Change.* Cambridge: Cambridge University Press.

Legendre, G., Miyata, Y., & Smolensky, P. (1990). Harmonic Grammar — A formal multi level connectionist theory of linguistic well formedness: Theoretical foundations. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (388–395). Cambridge, MA.

Legendre, G., Sorace, A., & Smolensky, P. (2006). The Optimality Theory–Harmonic Grammar connection. In P. Smolensky & G. Legendre (Eds.), *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar* (pp. 339–402). Cambridge: MIT Press.

Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.). (2013). *Ethnologue: Languages of the World, Seventeenth edition.* Dallas, TX: SIL International. Online version: http://www.ethnologue.com.

Lubowicz, A. (2002). Derived environment effects in Optimality Theory. *Lingua, 112,* 243–280.

Magri, G. (2012). Convergence of error-driven ranking algorithms. *Phonology, 29*(2), 213–269.

Martin, A. (2011). Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language*, *87*(4), 751–770.

Maye, J., Werker, J F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111.

McCarthy, J. J. (1998). Morpheme structure constraints and paradigm occultation. In M. C. Gruber, D. Higgins, K. Olsen, & T. Wysocki (Eds.), *Proceedings of the Chicago Linguistic Society 35. Vol. II: The panels* (pp. 123–150). Chicago, CLS.

McCarthy, J. J. (1999). Sympathy and phonological opacity. *Phonology, 16,* 331–399.

McCarthy, J. J. (2000). Harmonic Serialism and parallelism. In M. Hirotani (Ed.), *Proceedings of the North East Linguistics Society, 30.* Amherst, MA: GLSA.

McCarthy, J. J., & Prince, A. (1995). Faithfulness and reduplicative identity. In *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory.*

Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua, 122*(2), 145–163.

Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*(2), 338–352.

Minkova, D. (1993). On leapfrogging in historical phonology. In J. V. Marle (Ed.), *Historical Linguistics 1991: Papers from the 10th International Conference on Historical Linguistics, Amsterdam, 12-16 August 1991* (pp. 211–228)*.* Amsterdam: John Benjamins.

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

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.

Nagy, N., & Reynolds, B. (1997). Optimality theory and variable word-final deletion in Faetar. *Language Variation and Change, 9,* 37–55.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrik, & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior* (pp. 178–203)*.* Chicago: CLS.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). New York: Springer.

Ohala, J. (1997). Aerodynamics of phonology. *Proceedings of the Seoul International Conference on Linguistics* (pp. 92–97)*.* Seoul: Linguistics Society of Korea.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory exposure. *Cognition, 83*(1), B13–B23.

Pater, J. (2008). Gradual learning and convergence. *Linguistic Inquiry, 39*(2), 334–345.

Pater, J. (2009a). Review of Smolensky and Legendre (2006), *The Harmonic Mind. Phonology, 26,* 217–226.

Pater, J. (2009b). Weighted constraints in generative linguistics. *Cognitive Science, 33*(6), 999–1035.

Pater, J., Staubs, R., Jesney, K., & Smith, B. (2012). Learning probabilities over underlying representations. *Proceedings of the Twelfth Meeting of ACL-SIGMORPHON: Computational Research in Phonetics, Phonology, and Morphology* (pp. 62–71).

Pater, J., & Tessier, A-M. (2003). Phonotactic knowledge and the acquisition of alternations. In M-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1177–1180). Barcelona: Universitat Autónoma de Barcelona.

Pater, J., & Tessier, A.-M. (2005). Phonotactics and alternations: Testing the connection with artificial language learning. In K. Flack & S. Kawahara (Eds.), *University of Massachusetts Occasional Papers in Linguistics, 31,* 1–16.

Peperkamp, S., & Dupoux, E. (2002). Coping with phonological variation in early lexical acquisition. In I. Lasser, (Ed.), *The Process of Language Acquisition* (pp. 359–385). Frankfurt: Peter Lang.

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., Le Calvez, R., Nadal, J.-P., & Dupoux, E. (2006a). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition, 101,* B31–B41.

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

Potts, C., Pater, J., Jesney, K., Bhatt, R., & Becker, M. (2009). Harmonic Grammar with linear programming: from linguistic systems to linguistic typology. Ms., University of Massachusetts, Amherst.

Press, W. H., Flannery, B. P., Teukolsky, S A., & Vetterling, W. T. (1992). *Numerical recipes in C: The art of scientific computing.* Cambridge: Cambridge University Press.

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* (pp. 533–546). Somerville, MA: Cascadilla Press.

R Core Development Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Ross, K. (1996). *Floating phonotactics: Variability in infixation and reduplication of Tagalog loanwords*. M.A. thesis, UCLA.

Rubach, J. (1984). *Cyclic and Lexical Phonology: The Structure of Polish.* Dordrecht: Foris.

Rubach, J. (1993). *The Lexical Phonology of Slovak.* Oxford: Oxford University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606–621.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*(3), 484–494.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development, 1*(3-4), 289–316.

Singh, S., & Black, J. W. (1966). Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America, 39*(2), 372–387.

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). Somerville, MA: Cascadilla Press.

Skoruppa, K., & Peperkamp, S. (2011). Adaptation to novel accents: Feature-based learning of context-sensitive phonological regularities. *Cognitive Science, 35,* 348–366.

Smolensky, P. (1993). Harmony, markedness, and phonological activity. Handout at Rutgers Optimality Workshop – 1, October 23.

Smolensky, P. (1995). On the structure of *Con*, the constraint component of UG. Handout of talk at UCLA, April 7.

197

Smolensky, P. (1996). The initial state and "richness of the base" in Optimality Theory. Technical report, Johns Hopkins University.

Smolensky, P. (2006). Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature-domain markedness. In P. Smolensky & G. Legendre (Eds.), *The Harmonic Mind: From Neural Computation to Optimality Theoretic Grammar, Vol. 2: Linguistic and Philosophical Implications* (pp. 20–160). Cambridge: MIT Press.

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

Steriade, D. (1999). Alternatives to syllable-based accounts of consonantal phonotactics. In O. Fujimura, B. Joseph, & B. Palek (Eds.), *Proceedings of the 1998 Linguistics and Phonetics Conference* (pp. 205–245). Prague: The Karolinum Press.

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 313–334). Cambridge: Cambridge University Press.

Steriade, D. (2001/2008). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In S. Inkelas & K. Hanson (Eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky* (pp. 151–180). Cambridge: MIT Press. [Published in 2008. Originally circulated as ms. in 2001.]

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics, 39*(4), 505–513.

Tessier, A.-M. (2006). Testing for OO-Faithfulness in artificial phonological acquisition. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 619–630). Somerville, MA: Cascadilla Press.

Tessier, A.-M. (2012). Testing for OO-Faithfulness in the acquisition of consonant clusters. *Language Acquisition, 19*(2), 144–173.

Thatte, V. A. (2011). *Infant Phonotactic Learning.* Ph.D. dissertation, UCLA.

Trubetzkoy, N. S. (1939/1969). *Principles of Phonology.* (C. A. M. Baltaxe, Trans.). Los Angeles, Univ. of California Press. (Original work published 1939).

van der Hulst, H., & van de Weijer, J. (1995). Vowel harmony. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 495–534). Oxford: Blackwell.

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.

Werker, J. F., Shi, R., Desjardins, R., Pegg, J., Polka, L., & Patterson, M. (1998). Three methods for testing infant speech perception. In A. M. Slater (Ed.), *Perceptual development: Visual, auditory, and speech perception in infancy* (pp. 389–420). London: UCL Press.

White, K., Peperkamp, S., Kirk, C., & Morgan, J. (2008). Rapid acquisition of phonological alternations by infants. *Cognition, 107,* 238–265.

Wilson, C. (2001). Consonant cluster neutralization and targeted constraints. *Phonology, 18*(1), 147–197.

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

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy, 15*(4), 420–433.

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. *Language, 83,* 277–316.