Estimating historical probabilities of natural and unnatural processes

Gašper Beguš
begus@uw.edu

Abstract

This paper presents a technique for estimating the influences of channel bias on phonological typology. The technique based on statistical bootstrapping enables the estimation of Historical Probability, the probability that a synchronic alternation arises based on two diachronic factors — the number of sound changes required for an alternation to arise and their respective probabilities. We estimate Historical Probabilities of attested and unattested alternations, compare Historical Probabilities of alternations and perform inferential statistics on the comparison, and compare outputs of the diachronic model against the independently observed synchronic typology to evaluate the performance of the *channel bias* approach. The proposed technique also identifies mismatches in typological predictions of the *analytic bias* and *channel bias* approaches. By comparing these mismatches with the observed typology, this paper attempts to quantitatively evaluate the distinct contributions of the two influences on typology.

1 Introduction

Typological literature in phonology has long revolved around the question of which factors influence the observed typology. Two major lines of thought emerge in this discussion: the *analytic bias* approach and the *channel bias* approach (Moreton 2008, Yu 2013).[1] The *analytic bias* approach argues that the observed typology results primarily from differences in the learnability of phonological processes; the *channel bias* approach argues that the inherent directionality of sound changes based on phonetic precursors (articulatory and perceptual) shapes typology (for further discussion, see Hyman 1975, 2001; Greenberg 1978; Ohala 1981, 1983, 1993; Kiparsky 1995, 2006, 2008; Blevins 2004, 2006, 2007, 2008a,b; Wilson 2006; Hansson 2008; Moreton 2008; Hayes et al. 2009; Becker et al. 2011; Moreton and Pater 2012a,b; Moreton 2012; de Lacy and Kingston 2013; Garrett and Johnson 2013; Yu 2013; Hayes and White 2013; Cathcart 2015; Kirby and Sonderegger 2015; Greenwood 2016; i.a.).

Empirical evidence often supports both approaches equally well. Typologically frequent processes are often shown to directly result from phonologization of underlying articulatory or perceptual phonetic precursors (e.g. sound change in progress that results in a typologically common pattern), whereas rare or unattested processes lack such precursors, which supports the *channel bias* approach (cf. Hyman 1975; Greenberg 1978; Ohala 1981, 1983, 1993; Lindblom 1986; Barnes 2002; Blevins 2004, 2006, 2007, 2008a,b; see also Hansson 2008 and Garrett and Johnson 2013 for an overview of the literature). On the other hand, typologically rare processes are experimentally shown to be more difficult to learn, which supports the *analytic bias* approach (Kiparsky 1995, 2006, 2008; Wilson 2006; Hayes et al. 2009; Becker et al. 2011; de Lacy and Kingston 2013;

[1] Other names have been used for the two approaches, such as Evolutionary Phonology versus Amphichronic Phonology in Blevins (2004) and Kiparsky (2006, 2008).

Hayes and White 2013; White 2017; for an overview of the experimental *analytic bias* literature, see Moreton and Pater 2012a,b). This ambiguity of evidence poses the primary challenge in typological research. The stance of this paper is that both factors influence the typology (as has been argued by a mounting body of research recently; Hyman 2001, Myers 2002, Moreton 2008, Moreton and Pater 2012a,b, de Lacy and Kingston 2013). There is no doubt that all potential synchronic phonological processes need to be learnable and accommodated by the synchronic grammar and that all synchronic processes arise through some diachronic trajectory. The question that this paper addresses is whether observed typological distributions are influenced primarily by different degrees of the learnability of different processes or primarily by different diachronic trajectories that underlie different processes.

The role of phonological research is to quantitatively evaluate which aspects of typology are more likely to result from one factor or the other. Quantifying influences of the various factors that contribute to typology is a challenging task and several simplifying assumptions need to be made. Such an approach, however, brings advantages that are worth pursuing: a more transparent comparison between proposals. This paper proposes a technique based on which *channel bias* influences on typology can be estimated.

1.1 Analytic Bias

If typologically infrequent processes are experimentally shown to be more difficult to learn than typologically frequent processes (for an overview, see Moreton and Pater 2012a,b), a reasonable conclusion would be that typological observations result precisely from these differences in learnability. A challenge that the *analytic bias* approach faces is that artificial grammar learning experiments testing the learnability of typologically rare or nonexistent unnatural processes frequently fail to show learnability differences compared to typologically frequent natural processes when the structural complexity of the tested alternations is controlled for. Influences of *analytic bias* can be subdivided into *substantive bias* and *complexity bias* (Wilson 2006, Moreton 2008, Moreton and Pater 2012a,b). *Substantive bias* states that phonetically motivated processes are easier to learn than unmotivated (or unnatural). *Complexity bias*[2] states that alternations involving more conditioning features are more difficult to learn than simpler alternations (Moreton 2008). A survey of experimental literature on Analytic Bias in Moreton and Pater (2012a,b) shows that there exist consistent differences in experimental results testing the two biases. While *complexity bias* is consistently confirmed by the majority of studies surveyed, experimental outcomes of the *substantive bias* are mixed. Several studies that test the learning of unnatural alternations as defined in Section 2 found no effect of *substantive bias* (Pycha et al. 2003, Kuo 2009, Skoruppa and Peperkamp 2011, via Moreton and Pater 2012a,b; and more recently Seidl et al. 2007, Do et al. 2016, Glewwe 2017, Glewwe et al. 2018). A comparatively smaller subset of studies, however, do report positive results (Carpenter 2006, 2010; Wilson 2006).

L1 acquisition and L2 acquisition of word-final stops by speakers of L1s that ban obstruent codas is the only place where differences between the natural and unnatural pair of alternations are observed. Learners acquire word-final voiceless stops earlier than voiced stops and devoice voiced stops more frequently than they voice voiceless stops word-finally (overview in Broselow 2018; see also Clark and Bowerman 1986, Kong et al. 2012, and literature therein). It is likely, however, that this type of experiment tests differences in learning of more complex versus less complex articulations (Kong et al. 2012), and not the abstract phonological learning that is, for example, observed in artificial grammar learning experiments (e.g. where complex alternations

[2] *Complexity bias* has also been called Structural Bias (Moreton and Pater 2012a,b).

are more difficult to learn than simple alternations, which is independent of articulatory factors; Moreton and Pater 2012a,b). Articulation of segments that require more articulatory effort in a given position is expected to be learned less successfully: "[c]ross-language differences in the age of children's mastery of adult-like voiced stops are typically explained in terms of the relative difficulty of the laryngeal gestures for the language's voice onset time distributions" (Kong et al. 2012: 725). The very same mechanism is in fact responsible for final voicing within the *channel bias* approach: even adult L1 speakers with full contrast gradiently and passively devoice final voiced stops due to their greater articulatory complexity, which can result in a typologically common sound change that operates in an adult population (cf. Labov 1994). These L1 and L2 learning differences thus likely reflect differences in articulatory effort that should be modeled as a *channel bias* influence. It is in fact not trivial to show how differences in L1 articulatory learning would result in phonological typology (cf. Rafferty et al. 2013), given that children reproduce their input with a high degree of faithfulness past some developmental stage (e.g. at about 2–5 years for acquisition of the voicing contrast; see Kong et al. 2012).

1.2 Channel Bias

One of the objections against the *channel bias* approach to typology is that it fails to explain why some processes are unattested (Kiparsky 2006, 2008; de Lacy and Kingston 2013). Kiparsky (2006), for example, lists several diachronic trajectories that would lead to final voicing, yet final voicing is arguably not attested as a productive synchronic process. More generally, combinations of sound changes could conspire to yield a number of processes that are never attested as productive synchronic alternations. In the absence of a diachronic explanation, Kiparsky (2006) invokes grammatical constraints and learnability to explain these typological gaps.

Most of the current models of typology within the *channel bias* approach are indeed insufficient for explaining such typological gaps because they do not quantify the probability of the occurrence of sound changes or combinations of sound changes. The default explanation within the *channel bias* approach has long been a qualitative observation that common processes are frequent because they are produced by frequent sound changes or because they require less sound changes to arise (Blevins 2013:485, also Greenberg 1978:75–6). Such reasoning does not provide sufficient outputs for a quantitative comparison of different influences on phonological typology.

Despite these objections, mechanisms exist within the *channel bias* approach to derive typology beyond the simple statement that rare sound changes produce rare alternations. Based on a typological study of an unnatural process, post-nasal devoicing, Author (2017) argues that unnatural processes require at least three sound changes to arise (as opposed to at least two for unmotivated processes and at least one for natural processes, the so-called Minimal Sound Change Requirement), which explains the relative rarity of processes with different degrees of naturalness. To be sure, the idea that unmotivated processes are rare because they require a complex history is not new (Bell 1970, 1971; Greenberg 1978:75–6; Cathcart 2015; Morley 2015), but the Minimal Sound Change Requirement is, to the author's knowledge, the first formal proof that explains why unnatural processes are the least frequent (compared to natural or unmotivated processes, see Section 2). The Minimal Sound Change Requirement on its own, however, does not explain why some unnatural processes are attested, while others are not. To quantify the *channel bias* influences on typology further, the concept of the Minimal Sound Change Requirement should be combined with the estimation of probabilities of individual sound changes that are required for each synchronic alternation to arise.

Two models have thus far attempted to quantify probabilities of the occurrence of various primarily static phonotactic processes and explain the relative rarity of some processes based on diachronic factors. Bell (1970, 1971) and Greenberg (1978) propose a "state-process model". Their model operates with typological states (phonological, morphological, and syntactic) that can arise from other states, depending on transitional probabilities from one state to another and the rest probabilities of each state, and therefore is most suitable for modeling the probabilities of various phonotactic restrictions. The probability of each state is determined by the number of previous states from which it can arise and the transitional probabilities between the states. They propose a Markov chain model for determining the probabilities of each state. Modeling the probabilities of transitions (processes) in the instantiation of the model in Bell (1971) involves relative probabilities that only tangentially reflect the frequencies of the processes in the samples. While the main principles of Bell's (1970, 1971) and Greenberg's (1978) model are similar to what will be proposed in this paper, focusing the estimation of probabilities on sound changes, rather than on states, combined with using substantially more elaborate samples of sound changes in our model yields more accurate predictions (Section 4.3). A different model of calculating the probabilities of the combination of sound changes is offered by Cathcart (2015), who computes permutations of sound changes that lead to a certain process (in this case, final voicing) and compares that to permutations of all sound changes in a given survey to get an estimate of the probability of certain processes. Due to its design, however, Cathcart's (2015) model relies on representativeness of diachronic surveys for all sound changes, not only for the ones that are estimated (see also Section 3.2) and is computationally demanding, which makes the model difficult to implement. The models in Greenberg (1978) and Cathcart (2015) also do not take into consideration the crucial distinctions made in Author (2017): "the subdivision of unusual rules into unnatural versus unmotivated rules, paired with the proof that the former require at least three sound changes to arise".[3] The model proposed in this paper has a disadvantage that the trajectories of sound changes that lead to a certain alternation need to be identified manually (similar to Bell's (1971) and Greenberg's (1978) models), but this also means that samples of sound changes need to be representative only for the sound changes we are estimating. This paper argues that the Minimal Sound Change Requirement and a diachronic model of unnatural processes called the Blurring Process (Author 2017) facilitate the identification of trajectories that lead to unnatural processes, which consequently facilitates a quantitative model of typology within the *channel bias* approach.

1.3 Goals

The goal of this paper is to propose a quantitative method for estimating the influences of channel bias on phonological typology using a statistical method called bootstrapping (Efron 1979; Efron and Tibshirani 1994). The technique estimates the so-called Historical Probability, the probability that an alternation arises based on two diachronic factors — the number of sound changes required for an alternation to arise (the Minimal Sound Change Requirement; Section 2), and their respective probabilities, estimated from surveys of sound changes. The assumptions of the model are discussed in Section 3.3. This paper argues that with the technique, we can (i) estimate the Historical Probability of any alternation (Section 4.1), (ii) compare two alternations, attested or unattested, and perform statistical inferences on the comparison (Section 4.2), and (iii) compare outputs of the historical model with independently observed typology to evaluate the performance of the *channel bias* approach (Section 4.3). The paper also identifies mismatches in typological predictions between

[3]The model of automated reconstruction in Bouchard-Côté et al. (2013) estimates the probabilities of individual sound changes, but does not deal with combinations of sound changes. Other quantitative approaches to sound change (e.g. Kirby and Sonderegger 2013, 2015; Hruschka et al. 2015) do not directly deal with estimating the probabilities of sound changes that operate in combination, but computationally model the initiation and propagation of single sound changes.

the *analytic* and *channel bias* approaches (Sections 4.3 and 5). By testing these mismatched predictions against the observed typology, we can at least partially control for one factor when testing the other and vice versa, which consequently allows for quantitative evaluation of distinct contributions of the *analytic bias* and *channel bias* factors on phonological typology (Section 5).

2 Background

This paper adopts several diachronic concepts from Author (2017). First, this paper adopts the division of phonological processes into *natural*, *unmotivated*, and *unnatural*. Natural processes, such as final devoicing or post-nasal and intervocalic voicing, are cases of phonetically well-motivated universal phonetic tendencies. Unmotivated processes lack phonetic motivation, but do not operate against universal phonetic tendencies. Unnatural processes not only lack phonetic motivation, but also operate against universal phonetic tendencies.

This paper adopts two diachronic concepts for the derivation of typology within the *channel bias* approach that have been proposed in Author (2017): the Blurring Process and the Minimal Sound Change Requirement. Typological surveys of unnatural processes targeting the feature [±voice] conducted in Author (2017) and Author and Name (2017) identify thirteen languages in which PND has been reported either as a productive synchronic alternation or as a sound change. Based on this typological survey, a hypothesis about how unnatural processes arise diachronically is proposed: the Blurring Process which states that unnatural alternations arise through a combination of a specific set of three natural (phonetically motivated) sound changes: (i) a sound change that causes complementary distribution, (ii) a sound change that targets changed or unchanged segments in the complementary distribution, and (iii) a sound change that blurs the original complementary distribution (see Tables 1 and 2; Author 2017).

This allows us to maintain the long-held position that the operation of sound changes is limited to acoustically or perceptually motivated directions[4] (Garrett and Johnson 2013, Garrett 2014; for a discussion, see Author 1). This position has recently been challenged by Blust (2005), who lists a number of unnatural sound changes. If these unnatural sound changes result from combinations of natural sound changes (as argued by Author 2017 and Author and Name 2017), we can maintain the position that sound change is always phonetically motivated.

Unnatural alternations thus cannot arise from a single sound change. Author (1) additionally argues that unnatural segmental alternations cannot arise from two sound changes either. If a change from value + of a given feature ϕ_1 to value − is unnatural and therefore cannot result from a single sound change, some other feature (ϕ_2) has to change first, so that the change from $+\phi_1$ to $-\phi_1$ might be natural and motivated. To get the full unnatural process, however, that other feature (ϕ_2) has to change back to its original value. No such requirement exists for unmotivated processes: they can result from two sound changes. In sum, the *Minimal Sound Change Requirement* states that a minimum of three sound changes are required for an unnatural alternation to arise, a minimum of two sound changes for an unmotivated alternation, and a minimum of one sound change for a natural alternation (see Author 2017).

The proposed model does not account for unnatural processes that result from morphological changes. These are, however, almost always analyzed as morphologically conditioned and can be explained by non-phonological mechanisms.[5] The lack of modeling morphological changes is a short-

[4]In some cases, multiple directions of a sound change can potentially be motivated, although such cases are rare. For a discussion, see Author (Year).

[5]IVD in Sula (Bloyd 2015) likely results from a combination of morphological changes and is morphologically conditioned. The exact mechanisms of how IVD arises there are not straightforward and are beyond the scope of this

coming of the current proposal. Sound change is, however, substantially more unidirectional and sound change typology is substantially better understood compared to morphological or syntactic change, which is why a model that would include morphological sources of phonological processes would first require an elaborate typology of morphologically induced phonological processes.

3 Bootstrapping to estimate Historical Probabilities

If we disregard probabilities of individual sound changes, natural processes will thus be the most frequent, unmotivated less frequent, and unnatural the least frequent, because the former requires one sound change to arise, the second at least two, and the latter at least three (from Author 2017).

(1) *A scale of decreased probabilities (Author 2017)*
P(natural) > P(unmotivated) > P(unnatural)

Our goal is to propose a model that would quantify probabilities of natural, unmotivated, and unnatural processes further. We can combine the Minimal Sound Change Requirement with the assumption that the probabilities of sound changes influence the probabilities of synchronic alternations. The probability that an alternation arises based on diachronic factors depends on both the number of sound changes that are required for the alternation to arise and the probability of each individual sound change in the combination. Such probabilities are called *Historical Probabilities of Alternations* (P_χ).

A challenge in estimating the *channel bias* influences on typology is the possibility that the probabilities of each individual sound change estimated from diachronic surveys are influenced not only by the *channel bias* factors, but by learnability (*analytic bias*) as well (analytic bias-channel bias *conflation* problem henceforth; see Kiparsky 1995, 2008; Moreton 2008). That this is likely not the case for sound changes targeting the feature [±voice] is suggested by sound change typology: the natural sound changes PNV, IVV, and FV are frequent (Kümmel 2007), whereas their unnatural counterparts are never attested as individual *sound changes* (unlike unnatural synchronic alternations, which are rare but attested). Experiments that tested the learnability of these pairs of natural and unnatural processes, however, failed to find significant differences (Seidl et al. 2007, Do et al. 2016, Glewwe 2017; Glewwe et al. 2018). If learnability (*analytic bias*) factors played the primary role in determining probabilities of individual sound changes, we would expect unnatural sound changes to be as frequent as the natural ones.

It is still possible that learnability influences the probabilities of individual sound changes: learnability can promote or demote the likelihood of a phonetic precursor being phonologized (as argued by Moreton 2008; for criticism, see Yu 2011 and Kapatsinski 2011). Even if learnability indeed influences the probability of the phonologization of a process, its effects are likely minor, promoting or demoting phonologization, rather than affecting the probabilities of sound changes to the degree that we observe in the typology of natural and unnatural processes (Figure 3). That the effects are likely minor is suggested by learnability experiments: none of the artificial grammar learning experiments found a significant difference in learnability of natural and unnatural alternations (Seidl et al. 2007, Do et al. 2016, Glewwe 2017, Glewwe et al. 2018). Also, as is argued in Section 5, a phonologized sound change increases the complexity of a phonological system; the *complexity bias* would predict that learnability suppresses operation of sound change. These types of influences are not the primary interest of the paper (for disambiguation of the two influences, see Section 5).

paper.

Even if the probabilities of individual sound changes are crucially influenced by learnability (and therefore by *analytic bias*) and even if learnability causes a higher rate of operation of certain sound changes in combination, the requirement that more than one sound change needs to operate in a language for unmotivated alternations and unnatural alternations, respectively, to arise has to be independent of learnability, which means at least a portion of the estimated probabilities needs to be influenced by the *channel bias*.

Bootstrapping is a statistical technique within the frequentist framework for estimating sampling distribution (and consequently standard errors and confidence intervals for a statistic of interest) from a sample by random sampling with replacement (Efron 1979; Efron and Tibshirani 1994; Davison and Hinkley 1997). The computation is implemented in the statistical software R (R Core Team 2016) with the *boot* package (Canty and Ripley 2016, Davison and Hinkley 1997) using the functions *boot()* and *boot.ci()*. This paper also presents R code that implements the proposed technique (available in Appendix C).

3.1 The model

3.1.1 Individual sound changes

Probabilities of individual sound changes are estimated from a sample of successes (languages in a sample with a sound change S_i) and failures (languages in a sample without the sound change S_i), according to (2). If an alternation A_k requires only one sound change to arise and invariably occurs as a result of that change (i.e. A_k is natural), then we estimate its P_χ according to (2).

(2)

$$P_\chi(S_i) = \frac{\text{number of languages with sound change } S_i}{\text{number of languages surveyed}}$$

To estimate Historical Probability of a sound change using the bootstrapping technique, we sample with replacement from the sample of successes and failures (based on surveys of sound changes; see Tables 4 and 5) and calculate the statistic of interest. In our case, the statistic of interest is the probability according to (2). This is repeated 10,000 times (each sample being of the same length as the sample size). For example, when estimating Historical Probabilities of processes targeting feature [voice], the successes and sample sizes are taken from surveys of sound changes (for exact counts, see Section 4.1.2 and columns *Counts* and *Surveyed* in Tables 4 and 5 for successes and sample sizes, respectively). Sample sizes in our case range from 294 to 88, depending on the sound change (see Section 4.1.2). This repeated sampling with replacement yields a sampling distribution of Historical Probabilities: 10,000 data points. From this sampling distribution, standard error, bias, and 95% adjusted bootstrap (BC_a) confidence intervals that adjust for bias and skewness (Efron 1979, 1987) are computed.

3.1.2 Two or more sound changes

If an alternation A_k requires more than a single sound change, then the Historical Probability of A_k is estimated as a sum of the Historical Probabilities of each trajectory T_z that yields the alternation A_k, as shown in (3).

(3)

$$P_\chi(A_k) = P_\chi(T_1 \cup T_2 \cup \ldots \cup T_n)$$

A trajectory T_j denotes a combination of sound changes that yields an alternation A_k. In theory, there are an infinite number of trajectories that yield any given alternation, but for practical purposes, we estimate only the trajectory that involves the least number of sound changes. Historical Probabilities of trajectories that require more than three sound changes are assumed to be minor enough to be disregarded for practical purposes.

The Historical Probability of a trajectory T_j that requires more than a single sound change is estimated from the joint probability of the individual sound changes required for T_j, divided by the factorial of the number of sound changes in trajectory T_j if only one ordering results in the trajectory in question, as shown in (4).

(4)

$$P_\chi(T_j) = \frac{P_\chi(S_1 \cap S_2 \cap \ldots \cap S_n)}{n!}$$

Estimating the joint probability of individual sound changes ($P_\chi(S_1 \cap S_2 \cap \ldots \cap S_n)$) is not a trivial task. A number of assumptions are needed in order to compute this joint probability, the most important of which is the assumption that the occurrence of one sound change does not influence the probability of the following sound change. In other words, sound changes are treated as independent events. This is in fact a desirable assumption when modeling a purely diachronic approach to typology. As will be argued in Section 5, learnability does influence the probabilities of individual sound changes that operate in combination. Because these influences are in the domain of the *analytic bias* factor, however, they should not be modeled within the *channel bias* approach. For a discussion on the assumptions, including a discussion on the dependency of sound changes on phonemic inventories and learnability, see Section 3.3.

Historical Probability is a probability that a language L features an alternation A_k, regardless of the properties of L. In other words, we do not condition Historical Probabilities on languages that feature a certain property. The Historical Probability (P_χ) of the first individual sound change S_1 is thus estimated from the number of successes (languages with S_1) and the number of failures (languages without S_1) according to (2), regardless of the phonemic inventories of languages in the sample.

For example, if the target of the first sound change S_1 in a trajectory that results in an alternation A_k is a geminate stop, we estimate the Historical Probability of S_1 from the number of languages with the sound change S_1 divided by the number of all languages surveyed, including those that do not feature geminate stops. The Historical Probability of an alternation A_k that requires S_1 is simply the probability that the alternation A_k arises in a language L, regardless of whether it features stop geminates.

Once S_1 operates, however, we know that language L necessarily has the target/result/context of the sound change S_1. For this reason, we estimate the Historical Probability of the subsequent sound changes $P_\chi(S_2)$ by dividing the number of successes (languages with S_2) by the number of languages surveyed that feature the target/result/context of S_1 if these are also the target of S_2. The same is true for any subsequent sound change. Once we condition the probability of sound changes and estimate it from samples of sound changes given that they have the target/result/context of the previous sound change, we can treat the probabilities of individual sound changes as independent events under the *channel bias* approach and estimate P_χ from the product of the probabilities of individual sound changes (5).

(5)

$$P_\chi(T_j) = \frac{\prod_{i=1}^{n} P_\chi(S_i)}{n!}$$

To estimate standard errors and BC_a confidence intervals for a Historical Probability of A_k that requires more than a single sound change, the proposed technique samples with replacement from n individual binomial samples (one sample for each individual sound change, constructed as described above), computes the Historical Probability of each sound change (according to (2)), and then computes the product of the Historical Probabilities of each individual sound change divided by $n!$, according to (5). This process returns 10,000 bootstrap replicates of the Historical Probability of A_k, from which the standard errors and BC_a confidence intervals are computed.

3.1.3 Comparison

The proposed technique also allows for the estimation of the difference between the Historical Probabilities of two alternations, which consequently enables inferential statements on the comparison.

(6)

$$\Delta P_\chi(A_1, A_2) = P_\chi(A_1) - P_\chi(A_2)$$

The difference between the Historical Probabilities of two alternations (ΔP_χ) is estimated with a stratified non-parametric bootstrap, where P_χ of each individual alternation A_1 and A_2 is estimated as described in Sections 3.1.1 and 3.1.2 (depending on whether A_1 and A_2 require trajectories that require one or more sound changes). To compare two Historical Probabilities, we calculate the difference between $P_\chi(A_1)$ and $P_\chi(A_2)$, which returns 10,000 bootstrap replicates, from which the standard errors and BC_a confidence intervals are computed.

The proposed technique applied on a difference between two alternations enables a comparison of the two alternations with inferential statements. If the 95% BC_a confidence intervals of the difference both fall either below or above 0, then $P_\chi(A_1)$ and $P_\chi(A_2)$ are significantly different with $\alpha = 0.05$.

3.2 Sample

Samples used for estimating Historical Probabilities are created from counts of occurrences of sound changes in typological surveys. For a detailed description of how samples are created, see Section 4.1.2 and A. The technique is most accurate when typological surveys are large, well-balanced, and representative. Sound changes in a survey should always be evaluated with respect to the target of the change, its result, and its context. Sound change occurrences in a typological survey should be properly counted: if two or more daughter languages show the result of a sound change that operated at the proto-stage of the two languages, the sound change should be counted as a single event in the proto-language.

The most elaborate survey of sound changes currently available which we use in the paper is the survey of consonantal sound changes in Kümmel (2007). One major advantage of Kümmel's (2007) survey is that it includes language families with a well-reconstructed prehistory and a well-established subgrouping. This allows for a more accurate coding of the occurrence of a sound change, compared to competing surveys (e.g. the UniDia survey; Hamed and Flavier 2009). Sound changes are counted as single events if they operate at a proto-language stage. While it is sometimes difficult to reconstruct whether a sound change in two related languages operated at the proto-stage or independently in individual branches, especially for typologically frequent sound changes, the survey in Kümmel (2007) is the most comprehensive of all available surveys in this respect. While subgrouping and probabilities of sound change can be inferred through phylogenetic tree analysis (Hruschka et al. 2015), subgrouping in Kümmel's (2007) survey relies on historical methodology that includes information from both phonological as well as morphological and other higher level evidence.[6]

The survey in Kümmel (2007) includes approximately 294 languages and dialects of the Indo-European, Semitic, and Uralic language families. While the survey is limited to only three language families, the fact that it involves precisely those families that have well-established subgrouping, which allows for proper coding, compensates for the lack of representativeness. Results of the analysis presented in Section 4 are likely not crucially affected by the fact that many language families are excluded from the survey because frequencies and types of sound changes do not seem to be radically different across different language families (with recurrent sound changes appearing across all families; Blevins 2007; see also Section 4.3).

The only other comparable survey of sound changes known to the author is the UniDia database that surveys 10,349 sound changes from 302 languages (Hamed and Flavier 2009). The UniDia database is, however, less appropriate because it lacks elaborate diachronic subgroupings of languages. The survey appears to list changes from a proto-language to daughter languages irrespective of whether a change occurred at the proto-language stage or independently in the daughter languages. This means that changes that occurred at the proto-stage would be counted as independently occurring several times in the daughter languages, inflating their relative frequencies. This disqualifies the survey for our purposes. The database is not representative either, focusing primarily on the Bantu language family (83.5% of sound changes are from the Bantu family).

The proposed method offers advantages over Cathcart's (2015) proposal of estimating the probabilities of sound changes and their combinations. The requirement of sample representativeness is much weaker under the proposed approach. Cathcart's (2015) model crucially requires surveys of sound changes to be representative for all possible sound changes. Also, the model is based on the UniDia database, which is less appropriate compared to Kümmel's (2007) survey. Because the identification of historical trajectories that lead to an alternation is performed manually in the proposed model, surveys of sound changes that are used for calculations need not be representative for all possible sound changes, but only for those required for the alternation in question. In fact, elaborate surveys of sound changes can be constructed for each alternation in question even in the absence of a large and representative survey of sound changes.

3.3 Assumptions

As with any diachronic model, the proposed technique has to make some simplifying assumptions. In order to estimate the joint probability of two or more sound changes as a product of the Historical Probabilities of each individual sound change (see (5)), the model assumes that each sound change is an independent event. The proposed model *does* account for the dependency between sound changes where one sound change alters the target or context of the following sound change. Probabilities of sound changes are estimated based on their targets, results, and contexts (Section 3.2) and, crucially, from samples conditioned on the result of the previous sound change (Section 3.1.2). Two crucial assumptions of independence remain: that sound change is (i) independent of previous

[6] Additionally, phylogenetic tree analysis does not restrict the direction of sound change and would, for example, incorrectly analyze reported unnatural alternations as resulting from a single sound change.

sound changes when the dependence on targets, results, and contexts of the previous sound change is controlled for (Section 3.1.2) and (ii) independent of global phonemic properties of a language (those properties that do not immediately affect the conditions of sound changes in question).

The first assumption is not controversial when modeling typology within the *channel bias* approach. The proposed method aims to estimate only the *channel bias* influences on typology, which is why it has to assume that the probability of sound change is only determined by its frequency of operation evaluated on a diachronic and unconditioned level. In other words, the model assumes that sound change is blind to *analytic bias* factors such as the learnability of a process. While the probabilities of individual sound changes are modeled as independent of each other under the *channel bias* approach, it is likely that they are not independent: the operation of one change can influence the learnability of the resulting process, which consequently influences the operation of the following sound change. In fact, I will argue in Section 5 that probabilities of sound changes are indeed influenced by learnability factors and that a sound change that simplifies the learning of an alternation operates significantly more frequently than is predicted by only the *channel bias* factor. However, the diachronic model proposed here is designed to model only the *channel bias* contribution to the typology and should be blind to learnability, which means that the assumption of independence is desired for this purpose.

The second assumption of independence is more problematic: broader phonemic inventories can influence the probabilities of sound changes, especially for vocalic changes (e.g., due to the effects described in the Theory of Adaptive Dispersion, see Liljencrants and Lindblom 1972, Lindblom 1990), but also for consonantal changes. The proposed technique does not model the dependency of sound changes on those phonemic properties that do not immediately affect the targets, results, or contexts of the sound changes in question. The sample's representativeness should, however, cancel out potential dependencies. The sample of sound changes from which the Historical Probabilities are calculated includes languages with a diverse set of phonemic inventories (see Kümmel 2007). Additionally, we model Historical Probabilities of sound changes regardless of specific phoneme inventories. It is thus assumed that effects of the dependency between sound change and more distant phonemic properties does not crucially affect the probabilities of sound changes targeting the feature [±voice]. To the author's knowledge, no properties of phonemic inventories have been discovered that would affect the rate of the sound changes in question (e.g. intervocalic lenition, occlusion of fricatives, devoicing of stops).[7]

As already mentioned, identification of individual trajectories leading to an alternation A_k is performed manually in the current proposal. While this task is facilitated by the Blurring Process, which describes mechanisms for unnatural processes to arise, it is nevertheless possible that some trajectories that would potentially influence the final result are missing from the estimation. If we assume that the estimated trajectory T_j is indeed the most frequent trajectory leading to A_k and that potential alternative trajectories do not crucially influence the overall Historical Probability of an alternation, we can generalize the Historical Probability of that particular trajectory to the Historical Probability of the alternation. If such an assumption is not met, however, then the proposed technique estimates only the probability that an alternation A_k arises from a trajectory T_j.[8] This paper assumes that the estimated trajectories are the most frequent ones and that potential alternative trajectories do not crucially influence the results.

[7]The dependence of sound change on broader phonemic inventories is not modeled primarily because current surveys of sound changes are not sufficiently large and representative. In principle, the proposed technique could model this dependency by estimating the probabilities of sound changes from samples conditioned on some phonemic property of the surveyed languages.

[8]When more representative surveys become available, this assumption could be weakened by using Cathcart's (2015) permutation approach to identify trajectories for each alternation estimated with the proposed technique.

What is not accounted for in the model are the functional load of individual phonemes (Wedel 2012, Wedel et al. 2013, Hay et al. 2015) and other factors that could potentially influence probabilities of sound changes, such as lexical diffusion or lexical/morpheme frequency during the initial stages of sound change (Bybee 2002), language contact, and sociolinguistic factors. The model makes no assumptions about how sound change is initiated or spread. These factors can mostly be disregarded because the goal is to estimate the Historical Probability of alternation A_k operating in a language L with no conditional properties.

Finally, the proposed technique does not directly model the temporal dimension. In the absence of temporal information, we have to make some simplifying assumptions. These simplifying assumptions are not unique to the present proposal and are to some degree even desirable. The proposed technique estimates Historical Probabilities within a timeframe that approximates the average timeframe of the languages in the sample. The model also assumes that in order for a resulting alternation to be productive, all sound changes need to operate within one language L. While this might be too restrictive, it is, in fact, desirable to limit the timeframe in which sound changes and corresponding processes have to operate productively for the resulting alternation to be productive. For example, the combination of sound changes (the Blurring Process) that would result in PND in Yaghnobi operates over three languages and fails to result in a productive synchronic alternation. The model also assumes that once a sound change occurs in a language, it can reoccur. This is a closer approximation to reality than to assume that a sound change cannot operate in daughter languages once it has already operated in the parent language. In other words, sound changes in our model are birth-death events, a view that is substantiated by empirical evidence: sound change operates and then ceases to operate (Chen 1974), at which point it can occur again (e.g. on novel morphological or loanword material).

The Historical Probability of an unnatural alternation depends not only on sound changes that are required for the alternation to arise, but also on the probability that the opposite sound change (in our case, the natural sound change) will operate on the unnatural system and destroy the evidence for it. Currently, influences of the potential natural sound changes are not modeled because the Historical Probabilities of the natural sound changes (Table 8) are relatively similar for the processes estimated in this paper and we do not expect this additional factor to alter the results significantly. For other processes not estimated in this paper, including the probability of the natural sound change in the model might alter the outcomes significantly.

Most of the influences that are not directly modeled in the current proposal are at least partially accounted for by the fact that the sample size based on Kümmel's (2007) survey is relatively large and relatively representative. If the sample is representative, influences of various linguistic and non-linguistic factors will be reflected already in the sample and the results of the model will not be crucially affected. For practical purposes, these influences are disregarded. It is assumed that these effects are minor enough not to crucially affect the results, although such an assumption is currently difficult to evaluate.

4 Applications

4.1 Estimation of Historical Probabilities

4.1.1 Trajectories

The three natural alternations have obvious origins — the single natural sound changes PNV, IVV, and FD, respectively. For the unnatural alternations, we first identify sound changes in the Blurring Process (Section 2) that yield the alternation in question. If A > B / X is a natural sound change, then B > A / X is unnatural. Tables 1, 2, and 3 represent schematically (left column) how the unnatural B > A / X arises via the Blurring Cycle or the Blurring Chain (two subtypes of the Blurring Process; see Section 2, Author 2017, and Author and Name 2017). The actual sound changes that yield the unnatural alternation are identified in the right columns.

The origins of the unnatural alternations PND, IVD, and FV are well-established. A combination of the following three natural and well-motivated sound changes yields PND in all known cases: the fricativization of voiced stops in non-post-nasal position, the unconditioned devoicing of voiced stops, and the occlusion of voiced fricatives to stops (Table 1).[9]

Table 1: Blurring Cycle (schematic; left) yielding PND (right).

	Blurring Cycle	**PND**	**Schematic example**
			bamba
1.	B > C / ¬X	D > Z / [−nas]__	βamba
2.	B > A	D > T	βampa
3.	C > B	Z > D	bampa
Result	B > A / X	D > T / [+nas]__	

Author and Name (2017) argue that IVD results from three sound changes. Voiced stops fricativize intervocalically, voiced fricatives devoice, and voiceless fricatives get occluded to stops (see Table 2). The result is the unnatural intervocalic devoicing (D > T / V__V).

Table 2: Blurring Chain (schematic; left) yielding IVD (right).

	Blurring Chain	**IVD**	**Schematic example**
			baba
1.	B > C / X	D > Z / V__V	baβa
2.	C > D	Z > S	baɸa
3.	D > A	S > T	bapa
Result	B > A / X	D > T / V__V	

FV is arguably unattested both as a synchronic alternation and as a sound change[10] (Kiparsky 2006, Lipp 2016, Author 2017a, cf. Yu 2004, Rood 2016). A number of diachronic scenarios exist, however, that would yield FV and are identified in Kiparsky (2006). Most of the scenarios either include more than three sound changes or do not result in a phonological alternation but in a static phonotactic restriction instead (Section 4.1.2). One possible scenario that involves three sound changes and that would result in FV is Scenario 1[11] in Kiparsky (2006), which is used here for estimating the Historical Probability of FV. For the sound changes in Scenario 1 to result in synchronic alternations, we need to assume that geminate simplification first operated word-finally

[9] T represents voiceless stops, D voiced stops, S voiceless fricatives, and Z voiced fricatives.

[10] There is one possible case of FV that could count as a productive synchronic alternation — Lakota (Rood 2016). Currently, there are no acoustic studies of Lakota word-final stops. Since many reported cases of FV turned out to be cases of word-final unreleased or lax voiceless stops (Kiparsky 2006), I leave Lakota out of this discussion until acoustic data is available.

[11] Kiparsky's (2006) Scenario 2 also includes three sound changes, but the last sound change (apocope after a single consonant) is never attested in the UniDia database of sound changes (Hamed and Flavier 2009). Kümmel's 2007 survey does not include vocalic changes, which is why the UniDia database that surveys 10,349 sound changes from 302 languages is used. Because the last sound change is never attested in our surveys, I exclude Scenario 2 from the estimation of $P_\chi(FV)$.

and only later targeted other geminates. Without this assumption, the sound changes in Scenario 1 would result in a phonotactic restriction. The three sound changes operating to yield FV in this scenario are geminate simplification in word-final position, voicing of post-vocalic non-geminate stops, and unconditioned geminate simplification (see Table 3).

Table 3: Modified Blurring Chain (schematic; left) that would yield FV (right).

	Modified Blurring Cycle	**FV**	**Schematic example**
			pːapːapː
1.	C > B / X	Tː > T / __#	pːapːap
2.	B > A	T > D / V__	pːapːab
3.	C > B	Tː > T	papab
Result	B > A / X	T > D / __#	

4.1.2 Counts

Based on the trajectories identified here that result in natural and unnatural alternations, samples of sound changes based on which estimations of Historical Probabilities are performed are constructed from counts of occurrences and languages surveyed (from Kümmel's 2007 database). Sound change occurrences are counted from the number of languages that Kümmel (2007) lists for each sound change. To reduce counting a single sound change that operates at a proto-stage and is reflected in several daughter languages as independent events, sound changes with exact same outcome in closely related languages are counted as single events (as grouped together by Kümmel 2007). While it is possible that some dependencies still exist in the data, it is currently difficult to estimate them. We assume that potential dependencies do not crucially affect the results.

PNV as a sound change is reported in approximately 28 languages in Kümmel (2007). IVV is reported in approximately 38 languages (including post-vocalic voicing). FD is reported in approximately 24 languages (summarized in Table 4). For raw counts, see Sections A.1, A.2, and A.3. PNV, IVV, and FD that target a single series of stops are counted together with cases in which these sound changes target more than a single place of articulation.

Table 4: Counts of sound changes in Kümmel (2007) for natural alternations.

Alternation	**Sound change**	**Count**	**Surveyed**
PNV	T > D / N__	28	294
IVV	T > D / V__V	38	294
FD	D > T / __#	24	294

For the unnatural alternations that require more than a single sound change, counts are performed for each individual sound change in the corresponding Blurring Processes. If a sound change is reported to target a subset of the tree major places of articulation (labial, dorsal,velar) and not the entire set, the counts are multiplied by a coefficient that proportionally penalizes the counts. For example, counts of sound changes that target only two places of articulation are multiplied by $\frac{2}{3}$ in order to reduce the possibility of final estimated probabilities being inflated: if the first sound change targets two places of articulation and the second sound change targets the third place of articulation, such a combination would, for example, not result in an unnatural process.

The first sound change in the Blurring Chain that results in PND, the fricativization of voiced stops, is reported in approximately 66 languages. In 32 languages, the sound change is reported to target all three major places of articulation; in 11 languages the sound change targets two places of articulation, and in 23 languages one place of articulation. The final count is thus $32 + 11 \times \frac{2}{3} + 23 \times \frac{1}{3} \approx 47$. Instances of intervocalic and post-vocalic fricativization are included in the count (not only cases in which fricativization occurs in all but post-nasal position) because the result of such fricativization after the other two sound changes would be a system analyzed as PND as well.[12] The probability of the first sound change in the Blurring Cycle that results in PND is estimated based on the number of successes (languages in the survey with that sound change) and the total number of languages surveyed (294) without conditioning on the sample. The sample for estimating the probability of the first sound change is unconditioned because the Historical Probability of A_k is the probability that A_k arises in a language L, regardless of the properties of its phonemic inventory (see Section 3.1.2). Once the first sound change operates, however, we know that the language in question needs to have voiced stops in its inventory. The Historical Probability of the second sound change that targets voiced stops is therefore estimated from the number of successes (languages in the survey with that sound change) and the number of languages with voiced stops. The second sound change (D > T) is reported in approximately 15 ($\approx 13 + 1 \times \frac{2}{3} + 3 \times \frac{1}{3}$) languages (also counting cases of devoicing that are the result of chain shifts). Approximately 31 languages lack voiced stops in the survey in Kümmel (2007),[13] which means that P_χ is estimated based on $294 - 31 = 263$ languages surveyed. After the two sound changes operate, we also know that the language L has voiced fricatives. The Historical Probability of the last sound change is estimated based on the number of languages with occlusion of voiced fricatives and the number of languages surveyed with voiced fricatives (allophonic or phonemic). Approximately 217 languages in the survey have voiced (bi)labial, alveolar/dental, or velar non-strident fricatives,[14] according to Kümmel (2007). In approximately 17 ($\approx 1 + 5 \times \frac{2}{3} + 38 \times \frac{1}{3}$) languages, occlusion of fricatives is reported as a sound change. The counts for IVD are performed in the same manner as the counts for PND and are given in Table 5.

The Historical Probability of FV is estimated based on the one scenario in Kiparsky (2006) that would result in FV as an alternation. The scenarios that would lead to FV as a static phonotactic restriction and could involve fewer than three sound changes are excluded. There are three main reasons for why it is justified to distinguish alternations from static phonotactic restrictions in a diachronic model (Author 2017) despite the two phenomena likely being part of the same synchronic grammatical mechanisms (Prince and Smolensky 1993/2004, Hayes 2004, Pater and Tessier 2006). First, unnatural phonotactic restrictions that do not result from a Blurring Process provide considerably less reliable evidence for learners because the evidence is distributional rather than appearing within the same morphological unit across morphological boundaries. This means that the likelihood of a process not being acquired by the learners is considerably greater when it does not arise from a Blurring Process. Second, alternative analyses of data are often available in the case of phonotactic restrictions that do not result from a Blurring Process. Alternative explanations are not available in the case of alternations, where evidence for a process comes from

[12] An alternation that resulted from a combination of sound changes in which the first sound change targeted post-vocalic stops rather than non-post-nasal stops and the other two aforementioned sound changes have the same result as in the attested case of PND, and would be analyzed as PND with initial devoicing.

[13] One language has only /b/ in its inventory. The low number of inventories that lack voiced stops might be influenced by the areal that Kümmel (2007) surveys. Based on the PHOIBLE database (Moran et al. 2014), approximately 30% of inventories lack a phonemic labial voiced stop. For consistency purposes, we stay within Kümmel's (2007) survey with this acknowledgement.

[14] The labiodental voiced fricative /v/ is included in the count.

within the same morphological unit. Finally, typological surveys of phonotactic restrictions are considerably more difficult to establish (compared to typological surveys of alternations). In the absence of typological studies, it is difficult to evaluate predictions of the *channel bias* model for phonotactic restrictions. In fact, FV as a phonotactic restriction might not be as rare, with at least two potential phonological systems attested in which voiceless stops do not surface word-finally, but voiced stops do (Ho and some dialects of Spanish; see Author 2017). For a further discussion on the differences between phonotactic restrictions and alternations, see Author (2018).[15]

Counts of the sound changes that lead to FV as an alternation are as follows. In approximately three languages, word-final geminates are reported to simplify to singleton stops. (This sound change is necessary if we want the scenario to result in an unnatural alternation as opposed to a static phonotactic restriction.) Because this is the first in the series of changes and we do not condition P_χ on any property of language L, as before, the Historical Probability is estimated from the total number of languages surveyed. The second sound change, post-vocalic voicing of voiceless stops, is reported in approximately 23 languages (corrected for place of articulation). The intervocalic condition is excluded from the count, as voicing of intervocalic stops would not target final stops. Because all languages have voiceless stops, all 294 languages surveyed are included in the count for estimating the Historical Probability of the second sound change. Finally, simplification of geminates is reported in 21 languages. It is difficult to estimate how many languages in Kümmel (2007) allow geminate voiceless stops. While few languages have phonologically contrastive geminates, many more must allow sequences of two identical stops at morpheme boundaries (the so-called fake geminates; Oh and Redford 2012). To estimate the number of languages that allow such sequences, Greenberg's (1965) survey of consonantal clusters and Ryan's (to appear) survey of phonemic geminates are used. At least 30% of languages in Greenberg's (1965) survey of approximately 100 languages allow stop + stop final clusters. The number of languages in our sample that allow homorganic stop-stop sequences can be approximated from the proportion of languages that allow phonemic geminates and from the proportion of languages that allow sequences of stops. Languages that allow clusters of stops at morpheme boundaries should in principle allow clusters of homorganic stops: if geminate clusters were simplified, the sound change of simplification would of course be reported in our sample. The number is thus estimated at 88 (30% of 294 languages). That this estimate is accurate is suggested by a survey of phonemic geminates: Ryan (to appear) estimates that approximately 35% of 55 genealogically diverse languages surveyed have phonemic geminates.

Table 6 shows the Historical Probabilities with estimated 95% BC_a confidence intervals for the six natural and unnatural alternations discussed above. Figure 1 shows the distributions of bootstrap replicates for the Historical Probabilities (P_χ) of these natural and unnatural alternations. Table 6 and Figure 1 illustrate a substantial difference in Historical Probabilities between the natural and unnatural groups. The model thus predicts that the unnatural alternations (PND, IVD, and FV) will be substantially less frequent than their respective natural alternations (PNV, IVV, and FD).

4.2 Comparison of alternations

One of the advantages of the proposed model is that inferential statistics can be performed on the comparison between the Historical Probabilities of any two alternations. Significance testing is

[15] The scenario that potentially results in FV in Lakota is currently also excluded: the fricativization of voiceless stops word-finally, followed by post-vocalic voicing of fricatives and occlusion of fricatives to stops, would potentially result in FV. A preliminary estimation of this scenario shows that its Historical Probability would be very low because the first sound change is relatively rare (reported only once for one place of articulation in Kümmel 2007).

Table 5: Counts of sound changes in Kümmel (2007) for natural alternations.

Alternation	Sound change	Count	Surveyed
PND	D > Z / [−nas]/V__(V)	47	294
	D > T	15	263
	Z > D	17	216
IVD	D > Z / V__ (V)	42	294
	Z > S	5	216
	S > T	10	248
FV	T: > T / __#	3	294
	T > D / V__	23	294
	T: > T	21	≈88

Table 6: Estimated P_χ (in %) for natural and unnatural alternations with 95% BC_a confidence intervals. We also compute profile confidence intervals from an empty logistic regression for comparison. The highest difference between the confidence intervals is 0.5%, which suggests that the proposed model estimates CIs with high accuracy.

		95% BC_a CI		95% Profile CI	
A_k	P_χ	Lower	Upper	Lower	Upper
PNV	9.5	6.1	12.9	6.5	13.2
PND	0.01	0.006	0.02	—	
IVV	12.9	9.2	16.7	9.4	17.1
IVD	0.002	0.001	0.007	—	
FD	8.2	5.1	11.2	5.4	11.7
FV	0.003	0.001	0.01	—	

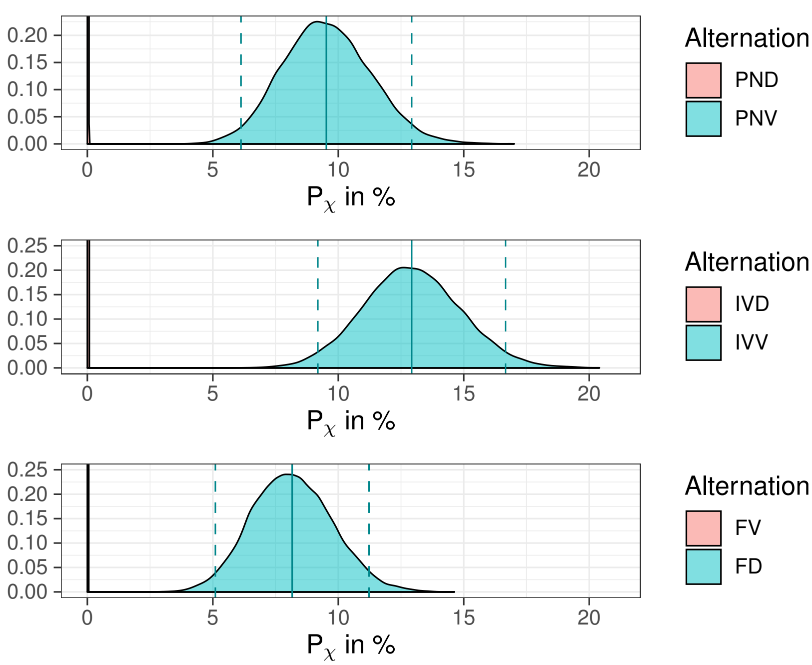

Figure 1: Bootstrap replicates for natural and unnatural alternations. The plots show the observed P_χ (solid line) and the 95% BC_a CI (dashed line) for natural alternations. The vast majority of bootstrap replicates for unnatural alternations fall outside the limits of the plot.

Table 7: Estimated ΔP_χ (in %) for natural-unnatural alternation pairs with 95% BC_a confidence intervals.

Alternation pair	Δ P_χ	95% BC_a CI Lower	95% BC_a CI Upper	
PNV vs. PND	9.5	6.5	13.3	*
IVV vs. IVD	12.9	9.5	17.0	*
FD vs. FV	8.2	5.4	11.9	*

performed by estimating a difference between the Historical Probabilities of two alternations (see Section 3.1.3). Samples of sound changes for estimating differences in Historical Probabilities are the same as the samples for estimating Historical Probabilities of individual alternations (Tables 4 and 5).

The Historical Probabilities of all three natural alternations in Figure 1 are significantly higher than the Historical Probabilities of their unnatural counterparts. Table 7 includes estimates and 95% BC_a confidence intervals of the difference in Historical Probabilities (ΔP_χ) for each natural-unnatural alternation pair. The calculations were performed as described in Section 3.1.3.

We can also compare alternations within the unnatural group. Figure 2 shows bootstrap replicates of the individual Historical Probabilities of the three unnatural alternations, PND, IVD, and FV. The figure shows that the Historical Probability of PND is higher compared to the Historical Probabilities of the other two unnatural alternations. By estimating the difference between two alternations, we can test, for example, whether P_χ(PND) and P_χ(IVD) or P_χ(PND) and P_χ(FV) are significantly different.

(7) a. $\Delta P_\chi(\text{PND}, \text{IVD}) = P_\chi(\text{PND}) - P_\chi(\text{IVD}) = 0.010\%\ [0.003\%, 0.02\%]$
b. $\Delta P_\chi(\text{PND}, \text{FV}) = P_\chi(\text{PND}) - P_\chi(\text{FV}) = 0.009\%\ [0.001\%, 0.02\%]$

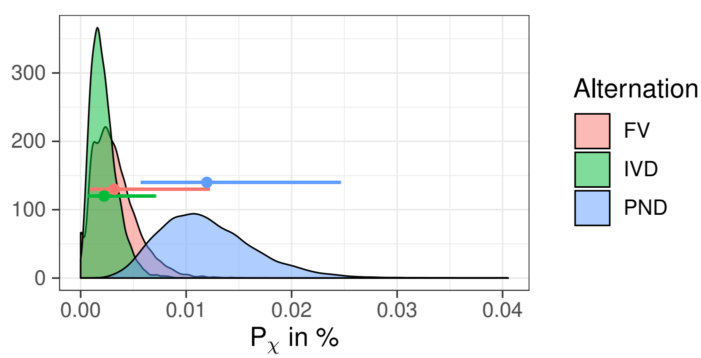

Figure 2: Bootstrap replicates for unnatural alternations with observed P_χ (colored dot) and 95% BC_a confidence intervals (colored lines).

Because the 95% BC_a CIs of the difference in Historical Probability between PND and FV and PND and IVD lie above zero, it can be concluded that P_χ(PND) is significantly higher than P_χ(FV) and P_χ(IVD) (with $\alpha = 0.05$).

Certainly, the proposed technique makes some simplifying assumptions that introduce confounds to the estimation of Historical Probabilities (see Sections 3.2, 3.3, and 4.1). Because differences in the Historical Probabilities between unnatural alternations are considerably smaller than differences between natural-unnatural pairs (Figure 1), estimation of these differences is substantially more prone to be influenced by these confounds and therefore less reliable. Until more comprehensive surveys are available, however, the proposed model makes, to the author's knowledge, the most accurate approximations of Historical Probabilities of alternations, both for natural-unnatural alternation pairs as well as for alternations within the unnatural group.

4.3 Comparing P_χ to observed synchronic typology

4.3.1 Natural vs. unnatural processes

We can evaluate the model's predictions by comparing Historical Probabilities with independently observed typology of *synchronic* alternations. Estimation of synchronic typological probabilities faces even more difficulties and problematic assumptions than estimation of Historical Probabilities. The presence of an alternation that results from a sound change in two related languages cannot be counted as independent, although it is often treated as such in synchronic typological surveys. Moreover, language contact and linguistic areas likely influence observed synchronic typology to a greater degree compared to the typology of sound changes, although this observation would need a more elaborate evaluation.

For all these reasons, a comparison of Historical Probabilities and observed synchronic typology can only be qualitative at this point, especially until more comprehensive and well-balanced surveys are available. Nevertheless, the estimated Historical Probabilities match the observed synchronic typology relatively well and, to the author's knowledge, better than alternative approaches (see Section 1.2). Table 8 compares Historical and observed synchronic probabilities. Historical Probabilities (P_χ) are estimated as described above (see Section 4.1 and Table 6). The synchronic typology is estimated with a non-parametric bootstrap technique in the same way as described in Section 3.1.1, except that the estimation is based on the number of languages in a sample with a synchronic alternation and the number of languages in a sample without the synchronic alternation. To be sure, synchronic typology is estimated from surveys of synchronic alternations, rather than from diachronic surveys of sound changes. The survey used for estimating synchronic typology is the P-base database (Mielke 2019), one of the largest surveys of synchronic phonological processes. Post-nasal voicing is attested in 28 languages, intervocalic voicing in 51 languages, and final devoicing in 31 languages. All three alternations are counted even if they target only one place of articulation. The P-base database surveys altogether 629 languages. Languages based on which the count is performed are given in Section B. Both the historical sample (Kümmel 2007) and the synchronic sample are not constructed specifically for the purpose of establishing typology of processes that target feature [voice], which makes them less prone to biases.

The following estimations of the synchronic typology of unnatural alternations can be computed (summarized in Table 8) based on surveys of unnatural processes in Author (2017) and Author and Name (2017). PND has been confirmed as a fully productive synchronic alternation in two related languages (Tswana and Shekgalagari) and as a morphophonological alternation in a few others (Buginese, Nasioi; see Author 2017). For the purpose of comparison, only fully productive alternations are counted in the synchronic typology. Because Tswana and Shekgalagari are closely related, PND here is counted as a single occurrence. IVD is attested only once as a morphologically conditioned synchronic process (Bloyd 2015), although detailed descriptions are lacking. FV is, to the author's knowledge, not attested as a productive phonological alternation in any language, which is why its synchronic typological probability is estimated below $P(\frac{1}{600})$.[16] An approximate estimate of languages surveyed in these surveys of unnatural alternations is 600.

[16] If we counted the best candidate for FV, Lakota, as featuring fully productive unnatural alternations (Rood 2016), the typological probabilities of FV would be estimated at $P(\frac{1}{600}) = 0.17\%$.

Table 8: A comparison of Historical Probabilities (P_χ) and observed synchronic typology (Typol.) with 95% BC_a CIs for natural and unnatural processes.

		95% BC_a CI			95% BC_a CI	
A_k	P_χ	Lower	Upper	Typol.	Lower	Upper
PNV	9.50	6.10	12.90	4.5	2.9	6.2
PND	0.01	0.006	0.02	0.5	0.0	1.2
IVV	12.9	9.2	16.7	8.1	6.0	10.2
IVD	0.002	0.001	0.007	0.2	0.0	0.5
FD	8.2	5.1	11.2	4.9	3.3	6.7
FV	0.003	0.001	0.01	0.0	0.0	0.0

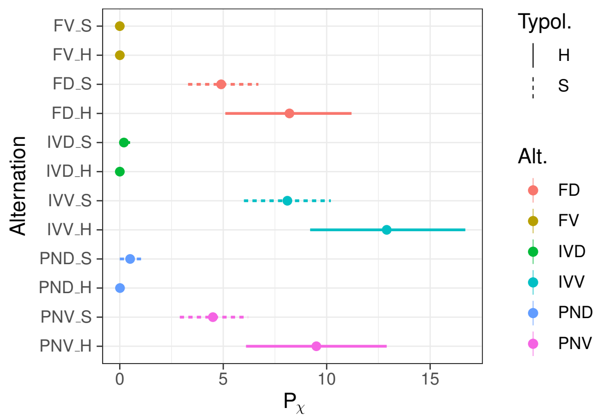

Figure 3: Observed Historical (H, solid line) and synchronic (S, dashed line) probabilities (in %) with 95% BC_a CIs from Table 8.

Table 8 and the corresponding plot of estimated Historical and synchronic probabilities with 95% BC_a CIs in Figure 3 suggest that the model correctly predicts natural alternations to be considerably more frequent than their unnatural alternations. Historical Probabilities and observed synchronic typology also match to the degree that the 95% BC_a confidence intervals of both Historical and synchronic typological probabilities always overlap for all five processes compared. It needs to be stressed here that for unnatural processes, the comparison of Historical Probabilities and observed synchronic typology is completely independent. In other words, the model estimates the probability of a combination of three sound changes, none of which are, by themselves, related to the unnatural synchronic alternation, from which synchronic typological probabilities are estimated.

4.3.2 Within the unnatural group

Inferential statements and predictions can be compared against the observed typology not only across natural-unnatural alternation pairs, but also within the unnatural group.

The model predicts PND to be significantly more frequent than FV and IVD (see (7) and Figure 2). Based on a survey in Author (2017) that aims to collect all reported cases of PND, the Blurring Process that leads to PND is attested in thirteen languages. In at least two languages, the Blurring Process results in a productive synchronic alternation, but it is likely that other languages, such as Buginese and Nasioi, feature the process as a productive alternation as well. IVD results from a Blurring Process in two languages. In one language, the Blurring Process results in a gradient phonotactic restriction (according to the survey in Author and Name 2017). In one additional case, IVD is attested as a morphologically conditioned synchronic alternation (Bloyd 2015). Finally, FV is, to the author's knowledge, never attested as a combination of sound changes or as a synchronic phonological alternation (for reasons why Lezgian and other cases are not analyzed as featuring FV, see Kiparsky 2006, Lipp 2016, and Author 2018). Although the comparison is currently qualitative, the typology suggests that PND is indeed more frequent than FV and IVD, as predicted by the model.

As already mentioned, alternations are distinguished from static phonotactics for the purposes of diachronic modeling (Section 4.1). Kiparsky (2006) lists a number of scenarios that would result in a static phonotactic restriction against voiceless stops with voiced stops surfacing word-finally. It is possible that some of his scenarios that require fewer than three sound changes would result in a productive unnatural phonotactic restriction, although this is less likely than in the case of unnatural alternations. It is also possible that *channel bias* alone cannot explain the relative rarity of FV and other unnatural phonotactic restrictions (as opposed to alternations). A further study is required to answer this question. A preliminary survey, however, suggests that FV as a static phonotactic restriction (as opposed to an alternation) might not be so rare, which would be expected under the proposed approach. Two languages might qualify as featuring this process: Ho and some varieties of Spanish (see Section 4.1 and Author 2017).

Kiparsky (2006, 2008) and others (de Lacy and Kingston 2013) claim that the *channel bias* approach to typology fails to explain why some processes, such as FV, are non-existent. The proposed model offers a potential solution to this problem. As was argued above, the model predicts that some unnatural processes will be significantly less frequent than others (or even unattested). FV is, for example, predicted to be very rare or possibly unattested, whereas PND is predicted to be significantly more frequent. This prediction seems to match the observed synchronic typology.

5 Implications

The proposed technique also helps identify mismatches in predictions between the *analytic bias* and *channel bias* approaches to typology. This section compares the outputs of the model to the predictions of the *analytic bias* approach to identify mismatches that help quantitatively estimate the influences of the two factors on typology.

If two typologically unequal alternations show no learnability differences, but have significantly different Historical Probabilities, it is reasonable to assume that the differences in the observed typology between the two alternations is influenced by the *channel bias* factor. On the other hand, if two typologically unequal alternations have equal Historical Probabilities and show differences in learnability, it is reasonable to assume that the differences result from the *analytic bias* factor. In the case of the unnatural alternations PND, IVD, and FV, the proposed model suggests that the observed typology is primarily influenced by the *channel bias* factor. The typology is predicted with relatively high accuracy (Section 4.3.1 and Figure 3), whereas learning experiments found no differences between the natural and unnatural alternations for any of the three pairs (Seidl et al. 2007, Do et al. 2016, Glewwe 2017, Glewwe et al. 2018).

The proposed technique further enables identification of mismatches in the predictions of the *analytic bias* and *channel bias* approaches, especially with respect to the complexity of alternations and their typological attestedness. The model predicts not only that unnatural alternations will be rare (Section 4), but also that, all else being equal, complex alternations will be less frequent than simple alternations. The minimality principle (Donegan and Stampe 1979, Picard 1994, Author 2017, and Section 2), which is at least a strong tendency, states that sound change is a change in one feature (or the deletion/reordering of feature matrices) in a given environment. This means that featurally complex alternations that change more than a single feature need to arise from the phonologization of more than one sound change. Because the probability of a combination of two sound changes will be lower than the probability of one sound change, all else being equal, featurally complex alternations are predicted to be typologically less frequent within the *channel bias* approach. Exactly the same generalization is, however, also predicted by the *analytic bias* approach to typology: numerous studies have confirmed that featurally complex alternations are consistently underlearned compared to featurally simple alternations (*complexity bias*; Moreton and Pater 2012a,b).

There is a crucial mismatch in predictions between the *analytic bias* and *channel bias* approaches with respect to unnatural alternations. The *channel bias* approach predicts that the more sound changes an alternation requires, the lower the Historical Probability of that alternation, regardless of its complexity (see Table 9). In other words, the prediction that complex alternations will be rare is violable: if the three sound changes of a Blurring Process result in a simple unnatural alternation, it will still be predicted that the simpler alternation will be less frequent than an unmotivated complex alternation because the first requires three sound changes to arise and the latter only two (Section 2).

We can estimate the Historical Probabilities for each step in the Blurring Process that leads to unnatural alternations. Let us take as an example PND. The Historical Probabilities of each resulting alternation (after the first, second, and third sound changes of the Blurring Process operate) were estimated as described in Section 3.[17] Table 9 (column P_χ) illustrates that each additional sound change decreases the Historical Probability of the resulting alternation.

On the other hand, the *analytic bias* approach predicts that structurally more complex alternations will be typologically less frequent because they are more difficult to learn than structurally simple alternations. (*complexity bias* has been confirmed almost without exception in many studies; Moreton and Pater 2012a,b.) If we analyze each step in the Blurring Process in terms of synchronic complexity, the first two sound changes in the Blurring Process indeed increase the complexity of the resulting alternation,[18] but the third sound change decreases its complexity. The alternation Z → T / [+nas] manipulates two feature values, [±continuant] and [±voice]. The alternation D → T / [−nas] (PND) manipulates only [±voice]. From a phonological perspective, the first is more complex than the latter (Moreton et al. 2017).

Complexity bias thus predicts that the alternations that arise from the first and second sound changes in the Blurring Process will be increasingly rare, but predicts that the structurally simpler alternations resulting from the combination of all three sound changes will be comparatively more frequent than the complex alternation requiring only two sound changes. Let us call this prediction the analytic bias-channel bias *complexity mismatch*. The exact mechanisms of how the *complexity bias* might affect the typology require further research. It is possible that the complex alternation is less likely to be learned faithfully, or that speakers will simplify the alternation (by reducing the number of features), or even that the last sound change will be blocked (Kiparsky 2008). Since no learnability differences are observed for natural-unnatural alternation pairs, the *analytic bias* approach stating that unnatural processes are rare because they are more difficult to learn faces

[17] The probability of the initial stage before the first sound change operates is calculated simply as $1-P_{\chi_{1,2,3}}$, where $P_{\chi_{1,2,3}}$ is the sum of the Historical Probabilities of the first, first and second, and all three sound changes.

[18] The fact that the first two sound changes in the Blurring Process occur relatively frequently, despite increasing the complexity of the alternations, argues against the radical approach to the *analytic bias-channel bias* conflation problem that states that sound change probabilities are primarily influenced by learnability and hence that estimated *channel bias* influences are crucially conflated with *analytic bias* influences. If anything, *analytic bias* influences would militate against the first two sound changes operating in combination because the resulting alternations would be more difficult to learn. Because the Blurring Process does occur, it means that the driving force behind the sound changes in question operating are not crucially influenced by *analytic bias* (although *analytic bias* can of course still influence the relative frequencies of sound change).

Table 9: Mismatches in predictions (framed) between the Channel Bias approach (P_χ) and the *complexity bias* approach (P_{cplx}) for PND. The *Sound Change* column represents the three sound changes from which the unnatural process PND results, and the *Alternation* column represents the synchronic alternation after each of the three sound changes. The P_χ column gives the estimated probability of each alternation with 95% BC_a Lower and Upper CIs (Lo. and Up.). The *Features* column counts the number of features a learner has to learn for each synchronic alternation.

Sound change	Alternation	P_χ	Lo.	Up.	Features	P_χ	P_{cplx}
	No alternation	83.5	—	—	0		
D > Z / [−nas]__	D → Z / [−nas]__	16.0	11.9	20.1	1	↓	↓
D > T	Z → T / [+nas]__	0.5	0.3	0.8	2	↓	↓
Z > D	PND	0.01	0.006	0.02	1	↓	↑

problems deriving the relative rarity of unnatural alternations as opposed to natural alternations. To be sure, a theory of markedness can nevertheless predict different frequency of processes, but such predictions are problematic if no learnability evidence exists to support them. Predictions of the *substantive bias* have been consistently questioned by the literature (Moreton and Pater 2012a,b).

The mismatched predictions illustrated in Table 9 provide new information for disambiguating *analytic bias* and *channel biases*. The *analytic bias*-*channel bias* complexity mismatch can be directly evaluated against the observed typology: if unmotivated structurally complex alternations that require two sound changes are typologically more common than structurally simpler unnatural alternations, *channel bias* has to be the leading cause of this particular typological observation. If, on the other hand, structurally more complex unmotivated alternations that require two sound changes are typologically less frequent than what would be predicted by the *channel bias* approach compared to structurally simpler unnatural alternations, we have a strong case in favor of the *analytic bias* influence, and more precisely in favor of *complexity bias* within the *analytic bias* approach to typology.

In fact, typological observations suggest that the complex synchronic alternation Z → T / [+nas]__ that results from the first two sound changes in a Blurring Process might be attested less frequently than would be predicted by *channel bias*, suggesting that *complexity bias* influences this distribution. The Historical Probability of Z → T / [+nas]__ is significantly higher than the Historical Probability of PND. The difference is estimated at $\Delta P_\chi(Z \rightarrow T / [+nas]\_\_, PND) = 0.4\%,\ [0.2\%, 0.8\%]$. In other words, the Historical Probability of the alternation Z → T / [+nas]__ that arises through two sound changes is predicted to be approximately fifty times more frequent than the Historical Probability of PND (see Table 9). Surface synchronic typology, however, does not conform to this generalization.

A system in which post-nasal devoiced stops contrast with voiced fricatives elsewhere (a complex alternation that arises via the combination of two sound changes) is synchronically confirmed in Konyagi, Punu, Pedi,[19] Sie, and potentially Nasioi (Dickens 1984; Hyman 2001; Merrill 2014, 2016a,b; Santos 1996; Brown 2017).[20] Other languages are more difficult to classify because some of them appear to feature full PND only for a subset of places of articulation. While Z → T / [+nas]__ indeed appears to be more frequent than PND, the magnitude of the difference appears

[19]We count Kutswe and Pulana together with Pedi, because they are closely related. Even if we counted them spearately (Kutswe and Pulana as one langauge), the distribution is still significant ($p = 0.016$).

[20]Punu is a language that undergoes a different development from the one described in Section 2. In Punu, the resulting alternation is not PND but the complex alternation between voiced fricatives elsewhere and voiceless stops post-nasally. For a discussion, see Hyman (2001).

to be smaller than predicted by the *channel bias*.

Even more intriguing is the high frequency at which the third sound change in the Blurring Process, occlusion of voiced fricatives to stops (Z > D), operates on synchronic systems that feature the alternation Z → T / [+nas]__ (after the first two changes in the Blurring Process). The Historical Probability of the third sound change in the Blurring Cycle that leads to PND, occlusion of voiced fricatives for languages that have voiced fricatives in the system, estimated independently of the Blurring Process (i.e. estimated from an unconditioned diachronic sample) is $\mathrm{P}_{\chi}(Z > D) = 20.4\%,\ [14.8\%, 25.5\%]$ (for languages that have voiced fricatives). Of the languages in the survey in Author (2017) that undergo the first two sound changes in the Blurring Process, which leads to PND, six languages (out of ten, or approximately 60%)[21] feature occlusion of stops for at least one place of articulation or in at least one position in the word. If we count only cases in which the occlusion of fricatives targets more than two places of articulation, only Tswana, Shekgalagari, Makuwa, and Murik would count. It does appear, however, that the occlusion of voiced fricatives in a synchronic system that undergoes the first two sound changes of the Blurring Cycle is more frequent than the model predicts for the occlusion of voiced fricatives in general.

To test the hypothesis that the last sound change operates with higher frequency than would be predicted by only the *channel bias* approach, we can compare the unconditioned Historical Probability of the occlusion of fricatives with the Historical Probability of the occlusion of fricatives in those languages that have already undergone the first two sound changes in the Blurring Cycle that lead to PND. In other words, we compare the probability of the occlusion of fricatives regardless of whether it simplifies the alternation (assuming only the *channel bias* influences) with the probability of the occlusion of fricatives operating in the Blurring Process, where it simplifies the alternation and consequently its learnability. Because both of these estimations involve a single sound change and because the second sample is small (eight observations), the significance of the difference is tested using Fisher's exact test. Counts for the unconditioned Historical Probability of the occlusion of fricatives is based on the survey of sound changes in Kümmel (2007). 44 languages with voiced fricatives (out of 216 surveyed) undergo the occlusion of voiced fricatives. As already mentioned, under the less conservative count, six out of ten languages with occlusion and devoicing of voiced fricatives show occlusion for at least one place of articulation or for at least one context (word-initially in Nasioi).[22] The difference between the two counts is statistically significant ($p = 0.009$, Fisher's Exact Test). This means that the last sound change in the Blurring Process that decreases the complexity of the resulting alternation operates at significantly higher rates than would be predicted if we only assumed *channel bias* influences.[23]

This suggests that the high occurrence of the third sound change in the Blurring Process (in the case of PND, the occlusion of fricatives) is likely an influence of *complexity bias* within the *analytic bias* approach. While *analytic bias* likely does not crucially influence the probabilities of the first two sound changes in the Blurring Process in the direction that interests us because they increase complexity and therefore would be predicted to reduce learnability (just as is predicted by the *channel bias* approach), it is likely that the occurrence of the third sound change, and therefore the lower probability of the more complex unmotivated alternation, is influenced precisely by *complexity bias*. The paper identifies and describes one such instance; investigation of further such cases should

[21]PND occurrences in Tswana, Shekgalagari, and Makhuwa are counted as only one occurrence. South Italian dialects that devoice affricates are not counted. I also exclude Mpongwe from the count because of the limited description and marginal status of PND there. I include Pedi which features Z → T / [+nas] (Dickens 1984) and Sie based on counts of the synchronic database of phonological rules in Mielke (2019).

[22]Cases with variation are counted as involving the sound change.

[23]This is exactly the opposite of what is proposed by Kiparsky (2008), who claims that the sound change that would result in an unnatural alternation would get blocked by the grammar.

yield a better understanding on how learnability and sound change frequency interact.

How exactly the analytic bias affects the increased frequency of the last sound change is subject to further research that includes experimental testing. Author (2018) proposes a potential mechanism: voiced fricatives often include variation with voiced stops. The stop variants can be relatively infrequent, but a bias that favors simple alternations might skew the initial distribution into a system in which voiced stops become increasingly more frequent. Author (2018) presents experimental evidence in support of this hypothesis.

6 Conclusion

This paper proposes a technique for estimating channel bias influences on phonological typology using the statistical technique bootstrapping. We estimate Historical Probabilities of alternations that are based on two diachronic factors: the number of sound changes required for an alternation to arise and their respective probabilities. The paper provides a detailed description of the statistical model and discusses its assumptions, properties of the sample, and implementation. This paper also includes functions in the statistical software R (R Core Team 2016) for performing the analysis.

Several applications of the model are presented. The model (i) estimates the Historical Probability of any synchronic alternation, both attested and unattested, (ii) compares the Historical Probabilities of two alternations and performs inferential tests on the comparison, and (iii) compares the Historical Probabilities to independently observed synchronic typology to evaluate the *channel bias* influences on typology. Finally, we identify mismatches in predictions between the *analytic bias* and *channel bias* approaches, which yields new insights into the discussion of different influences on phonological typology.

Both *analytic bias* and *channel bias* approaches predict that complex alternations will be less frequent than simple alternations, but within the *channel bias* approach this prediction can be violated in the case of unnatural vs. unmotivated alternations. This paper suggests that occlusion of voiced fricatives operates significantly more frequently as the last sound change in the Blurring Process, where it simplifies an alternation, compared to its operation in an unconditioned sample (where no simplification occurs). In other words, the sound change that simplifies a complex alternation operates significantly more frequently than it would as predicted by only the *channel bias* approach, suggesting that the *analytic bias* factor is responsible for the typological distribution.

The results suggest that the typological difference between natural and unnatural alternations targeting the feature [±voice] is primarily due to *channel bias*, but that the relatively low frequency of complex alternations and the higher rate of the operation of sound changes that simplify an alternation are due to *analytic bias*.

These conclusions have direct theoretical implications. Synchronic grammar should ideally derive all observed patterns and at the same time exclude impossible processes. Typological observations often prompt adjustments in grammar design. The proposed framework suggests that some typological gaps are historical accidents that need not be encoded in synchronic grammars, and quantifies these gaps. On the other hand, this paper also suggests that some typological observations, such as the avoidance of complex alternations, cannot be explained only within the *channel bias* approach and that these preferences should indeed be encoded in synchronic grammar. Estimation of the *channel bias* and *analytic bias* influences should thus be performed on further alternations in order to gain a better understanding of which observations result from constraints in synchronic grammar and which from diachronic development.

References

Barnes, Jonathan. 2002. Positional neutralization: a phonologization approach to typological patterns. Ph.D. dissertation, University of California, Berkeley.

Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1): 84-125.

Bell, Alan. 1970. A state-process approach to syllabicity and syllabic structure. Ph.D. Dissertation, Stanford University.

Bell, Alan. 1971. Some patterns of the occurrence and formation of syllabic structure. *Working Papers on Language Universals* 6: 23-138.

Blevins, Juliette. 2004. *Evolutionary Phonology.* Cambridge: Cambridge University Press.

Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32: 117-165.

Blevins, Juliette. 2007. The importance of typology in explaining recurrent sound patterns. *Linguistic Typology* 11:107-113.

Blevins, Juliette. 2008a. Natural and unnatural sound patterns: a pocket field guide. In *Naturalness and Iconicity in Language*, edited by Klaas Willems and Ludovic De Cuypere. 121-148. Amsterdam: Benjamins.

Blevins, Juliette. 2008b. Consonant epenthesis: natural and unnatural histories. In *Language Universals and Language Change*, edited by Jeff Good. 79-107. Oxford: Oxford University Press.

Blevins, Juliette. 2013. Evolutionary Phonology: A holistic approach to sound change typology. In *Handbook of Historical Phonology*, edited by Patrick Honeybone and Joseph Salmons. Oxford: Oxford University Press.

Bloyd, Tobias. 2015. Toward a Phonological Reconstruction of Proto-Sula. *University of Hawai'i at Mānoa: Working Papers in Linguistics* 46(8): 1-23.

Bloyd, Tobias. 2017. Synchronic intervocalic fortition in Sula: a counter-universal. Talk at the *91st Annual Meeting of the Linguistic Society of America*, January 5-8, 2017, Austin, TX.

Blust, Robert. 2005. Must sound change be linguistically motivated? *Diachronica* 22(2): 219-269.

Blust, Robert. 2013. *The Austronesian Languages.* Canberra: Asia-Pacific Linguistics.

Bouchard-Côté, Alexandre, David Hall, Thomas L. Griffiths, and Dan Klein. 2013. Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences*, 110(11): 4224-4229.

Broselow, 2018. Laryngeal contrasts in second language phonology. In *Phonological typology*, edited by Larry M. Hyman and Frans Plank. 312-340. Berlin: De Gruyter.

Brown, Jason. 2017. Postnasal devoicing in Nasioi. *Oceanic Linguistics* 56(1): 267-277.

Buckley, Eugene. 2000. On the naturalness of unnatural rules. In *Proceedings from the Second Workshop on American Indigenous Languages. UCSB Working Papers in Linguistics* 9: 1-14.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261-290.

Canty, Angelo and Brian Ripley. 2016. *boot: Bootstrap R (S-Plus) Functions.* R package version 1.3-18.

Carpenter, Angela C. 2006. Acquisition of a natural versus an unnatural stress system. Ph. D. Dissertation, University of Massachusetts, Amherst.

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

Cathcart, Chundra A. 2015. A probabilistic model of Evolutionary Phonology. *Proceedings of the Forty-Fifth Annual Meeting of the North East Linguistic Society.* Vol. 1, edited by Thuy Bui and Deniz Özyıldız. 145-150. Amherst, MA: GLSA.

Chen, Matthew Y. 1974. Natural phonology from the diachronic vantage point. In *Papers from the Parasession on Natural Phonology*, edited by A. Bruck, R. A. Fox, and M. W. LaGaly. 21-29. Chicago, IL: Chicago Linguistic Society.

Clark, Eve V., and Melissa Bowerman. 1986. On the acquisition of final voiced stops. In *The Fergusonian Impact: In Honor of Charles A. Ferguson on the Occasion of his 65th Birthday.*, edited by J. A. Fishman. 51-68. Volume 1: From Phonology to Society. Berlin: De Gruyter.

Coetzee, Andries W. and Rigardt Pretorius. 2010. Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38: 404-421.

Davison, A. C. and D. V. Hinkley. 1997. *Bootstrap Methods and Their Applications.* Cambridge: Cambridge University Press.

de Lacy, Paul and John Kingston. 2013. Synchronic explanation. *Natural Language and Linguistic Theory.* 31(2): 287-355.

Dickens, Patrick J. 1984. The history of so-called strengthening in Tswana. *Journal of African Languages and Linguistics* 6:97-125.

Do, Youngah, Elizabeth Zsiga, and Jonathan Havenhill. 2016. Naturalness and frequency in implicit phonological learning. Talk presented at the *90th Annual Meeting of the Linguistic Society of America*, Washington, DC, January 7-10, 2016.

Donegan, Patricia J. and David Stampe. 1979. The study of Natural Phonology. In *Current Approaches to Phonological Theory*, edited by Daniel Dinnsen. 126-173. Bloomington: Indiana University Press.

Efron, Bradley. 1979. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7(1): 1-26.

Efron, Bradley. 1987. Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82(397): 171–185. doi:10.2307/2289144.

Efron, Bradley, and Robert J. Tibshirani. 1994. *An introduction to the bootstrap*. CRC press.

Garrett, Andrew. 2014. Sound change. In *The Routledge Handbook of Historical Linguistics*, edited by Claire Bowern and Bethwyn Evans. 227-248. New York: Routledge.

Garrett, Andrew and Keith Johnson. 2013. Phonetic bias in sound change. In *Origins of Sound Change: Approaches to Phonologization*, edited by Alan C. L. Yu. 51-97. Oxford: Oxford University Press.

Glewwe, Eleanor. 2017. Substantive bias in phonotactic learning: Positional extension of an obstruent voicing contrast. Talk presented at the *53rd meeting of Chicago Linguistic Society*, Chicago, IL, May 25-27, 2017.

Glewwe, Eleanor, Jesse Zymet, Jacob Adams, Rachel Jacobson, Anthony Yates, Ann Zeng, Robert Daland. 2018. Substantive bias and word-final voiced obstruents: an artificial grammar learning study. Paper presented at the *92nd Annual Meeting of the Linguistic Society of America*. Salt Lake City, UT, January 4-7, 2018.

Greenberg, Joseph H. 1965. Some generalizations concerning initial and final consonant sequences. *Linguistics* 18: 5-34.

Greenberg, Joseph H. 1978. Diachrony, synchrony, and language universals. In *Universals of Human Language*, edited by Joseph H. Greenberg. Volume I: Method & Theory. 61-92. Stanford, CA: Stanford University Press.

Greenwood, Anna. 2016. An experimental investigation of phonetic naturalness. Ph.D. Dissertation, University of California, Santa Cruz.

Hamed, Mahé B. and Sébastien Flavier. 2009. UNIDIA: A database for deriving diachronic universals. In *Historical Linguistics 2007: Selected Papers from the 18th International Conference on Historical Linguistics, Montreal, 6–11 August 2007*, edited by Monique Dufresne, Fernande Dupuis, and Etleva Vocaj. 259-268. Amsterdam: John Benjamins.

Hansson, Gunnar. 2008. Diachronic explanations of sound patterns. *Language and Linguistics Compass* 2:859-893.

Haspelmath, Martin. 1993. *Lezgian grammar*. Berlin: Mouton de Gruyter.

Hay, Jennifer B., Janet B. Pierrehumbert, Abby J. Walker, and Patrick LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139: 83-91.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In *Constraints in Phonological Acquisition*, edited by R. Kager, J. Pater and W. Zonneveld. 153-208. Cambridge: Cambridge University Press.

Hayes, Bruce and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44(1): 45-75.

Hayes, Bruce, Péter Siptár, Kie Zuraw, and Zsuzsa Londe 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85(4): 822-863.

Hruschka, Daniel J., Simon Branford, Eric D. Smith, Jon Wilkins, Andrew Meade, Mark Pagel, and Tanmoy Bhattacharya. 2015. Detecting regular sound changes in linguistics as events of concerted evolution. *Current Biology* 25(1): 1-9.

Hyman, Larry M. 1975. *Phonology: Theory and Analysis.* New York: Holt, Rinehart & Winston.

Hyman, Larry M. 2001. The limits of phonetic determinism in phonology: *NC revisited. In *The Role of Speech Perception in Phonology*, edited by Elizabeth Hume and Keith Johnson. 141–186. San Diego, CA: Academic Press.

Hyman, Larry M. 2013. Enlarging the scope of phonologization. In *Origins of Sound Change: Approaches to Phonologization*, edited by Alan C. L. Yu. 3-28. Oxford: Oxford University Press.

Iverson, Gregory K. and Joseph C. Salmons. 2011. Final devoicing and final laryngeal neutralization. In *The Blackwell Companion to Phonology: Suprasegmental and Prosodic Phonology.* Volume 2, edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice. 1622-1643. Malden: Wiley-Blackwell.

Kapatsinski, Vsevolod. 2011. Modularity in the channel: The link between separability of features and learnability of dependencies between them. In *Proceedings of the XVIIth International Congress of Phonetic Sciences*, edited by W. S. Lee and E. Zee. 1022-1025. Hong Kong.

Kiparsky, Paul. 1995. The phonological basis of sound change. In *Handbook of Phonological Theory*, edited by John Goldsmith. 640-670. Oxford: Blackwell, 1995.

Kiparsky, Paul. 2006. Amphichronic program vs. Evolutionary Phonology. *Theoretical Linguistics* 32(2): 217-236.

Kiparsky, Paul. 2008. Universals constrain change, change results in typological generalizations. In *Linguistic universals and language change*, edited by Jeff Good. 23-53. Oxford: Oxford University Press.

Kirby, James and Morgan Sonderegger. 2013. A model of population dynamics applied to phonetic change. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, edited by Markus Knauff, Michael Pauen, Natalie Sebanz, and Ipke Wachsmuth. 776-781. Austin, TX: Cognitive Science Society.

Kirby, James and Morgan Sonderegger. 2015. Bias and population structure in the actuation of sound change. *arXiv*: 1507.04420.

Kong, Eun Jong, Mary E. Beckman, and Jan Edwards. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics* 40(6): 725-744.

Kümmel, Martin. 2007. *Konsonantenwandel.* Wiesbaden: Reichert.

Kuo, Li-Jen. 2009. The role of natural class features in the acquisition of phonotactic regularities. *Journal of psycholinguistic research* 38(2): 129-150.

Labov, William. 1994. *Principles of Linguistic Change.* Volumes 1 and 2. Oxford: Blackwell.

Liljencrants, Johan and Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48(4): 839-862.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In *Speech Production and Speech Modelling*, edited by William J. Hardcastle and Alain Marchai. 403-439. Dordrecht: Kluwer.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In *Experimental Phonology*, edited by John J. Ohala and Jeri J. Jaeger. 13-44. Orlando: Academic Press.

Lipp, Rainer. 2016. Final stops in Indo-European: their phonological classification as a key to the Proto-Indo-European root structure constraint. *Slovo a slovesnost* 77: 251-299.

Merrill, John. 2014. A historical account of the Fula and Sereer consonant mutation and noun class systems. Ms., University of California, Berkeley.

Merrill, John. 2016a. Consonant mutation and initial prominence: The historical loss of lexical contrastiveness. Talk presented at the *90th Annual Meeting of the Linguistic Society of America*, Washington, DC, January 7-10, 2016.

Merrill, John. 2016b. Konyagi post-nasal devoicing? Ms., University of California, Berkeley.

Mielke, Jeff. 2019. PBase: A database of phonological patterns. Retrieved May 12, 2020 from `/pbase.phon.chass.ncsu.edu`.

Moran, Steven, Daniel McCloy, and Richard Wright. 2014. *PHOIBLE Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at `phoible.org` on May 6, 2018.

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

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

Moreton, Elliott, and Joe Pater 2012a. Structure and substance in artificial-phonology learning. Part I, Structure. *Language and Linguistics Compass* 6 (11): 686-701.

Moreton, Elliott, and Joe Pater 2012b. Structure and substance in artificial-phonology learning. Part II, Substance. *Language and Linguistics Compass* 6 (11): 702-718.

Moreton, E., Pater, J. and Pertsova, K. (2017), Phonological Concept Learning. Cogn Sci, 41: 4-69. doi:10.1111/cogs.12319

Morley, Rebecca L. 2015. Can phonological universals be emergent? Modeling the space of sound change, lexical distribution, and hypothesis selection. *Language* 91(2): e40-e70.

Myers, Scott. 2002. *Gaps in factorial typology: the case of voicing in consonant clusters.* Ms., University of Texas, Austin. ROA.

Oh, Grace E. and Melissa A. Redford. 2012. The production and phonetic representation of fake geminates in English. *Journal of Phonetics* 40(1): 82-91. doi: 10.1016/j.wocn.2011.08.003.

Ohala, John J. 1981. The listener as a source of sound change. In *Papers from the Parasession on Language and Behavior*, edited by Carrie S. Masek, Robert A. Hendrick, and Mary Frances Miller. 178–203. Chicago: Chicago Linguistic Society.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, edited by Peter F. MacNeilage. 189-216. New York: Springer.

Ohala, John J. 1993. The phonetics of sound change. In *Historical Linguistics: Problems and Perspectives*, edited by Charles Jones. 237-278. London: Longman.

Ohala, John J. 2011. Accommodation to the aerodynamic voicing constraint and its phonological relevance. In *Proceedings of the 17th International Congress of Phonetic Sciences*. 64-67.

Pater, Joe and Anne-Michelle Tessier. 2006. L1 phonotactic knowledge and the L2 acquisition of alternations. In *Inquiries in Linguistic Development: In honor of Lydia White*, edited by Roumyana Slabakova, Silvina A. Montrul and Philippe Prévost. 115-131. Amsterdam: John Benjamins.

Picard, Marc. 1994. *Principles and Methods in Historical Phonology: From Proto-Algonkian to Arapaho.* Montreal: McGill-Queen's University Press.

Prince, Alan. and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell. First published in 1993, Tech. Rep. 2, Rutgers University Center for Cognitive Science.

Pycha, Anne, Pawel Nowak, Eurie Shin, and Ryan Shosted. 2003. Phonological rule-learing and its implications for a theory of vowel harmony. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, edited by Gina Garding and Mimu Tsujimura. 101-114. Somerville, MA: Cascadilla Press.

R Core Team 2016. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rafferty Anna N., Thomas L. Griffiths, and Marc Ettlinger. 2013. Greater learnability is not sufficient to produce cultural universals. *Cognition* 129(1): 70-87.

Rood, David. 2016. The phonology of Lakota voiced stops. In *Advances in the Study of Siouan Languages and Linguistics*, edited by Catherine Rudin and Bryan J. Gordon. 233-255. Berlin: Language Science Press.

Ryan, Kevin. To appear. *Prosodic weight: categories and continua.* Oxford University Press.

Santos, Rosine. 1996. *Le Mey: langue ouest-atlantique de Guinée.* Paris: Université Paris III.

Seidl, Amanda, Eugene Buckley, and Alejandrina Cristià. 2007. Complexity trumps naturalness. Talk presented at the *81st Annual Meeting of the Linguistic Society of America*, Anaheim, CA, January 4-7, 2007.

Skoruppa, Katrin and Sharon Peperkamp. 2011. Adaptation to novel accents: feature-based learning of context-sensitive phonological regularities. *Cognitive Science* 35: 348-366.

Wedel, Andrew. 2012. Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition* 4(4): 319-355.

Wedel, Andrew, Abby Kaplan, and Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2): 179-186,

Wilson, Colin. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30: 945-982.

Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80: 73–97.

Yu, Alan C. L. 2011. On measuring phonetic precursor robustness: a response to Moreton. *Phonology* 28(3): 491-518.

Yu, Alan C. L. 2013. Individual differences in socio- cognitive processing and sound change. In *Origins of Sound Change: Approaches to Phonologization*, edited by Alan C. L. Yu. 201-227. Oxford: Oxford University Press.

A Counts

A.1 *PNV*

Three places of articulation: Milian, Lycian, Karian, Lydian; North-West Middle Indo-Aryan; New Indo-Aryan Dom-Lom-Rom, Sindhi, Lahnda, Punjabi, Western Pahari; Common East Iranian; Late Old Persian, Common West Middle Iranian; Middle Armenian, Common Armenian; Common Albanian; Middle Greek; Common South Italian; Spanish (Pyrenees); Scottish Gaelic, Manx; East Saami; East Saami (Skolt, Patsjoki, Inari); Central Saami; Mari; Hungarian; Common Permic; Enets; North, Middle (dialectal), South Selkup
Two places of articulation: Proto-Irish; Milian, Lycian, Karian; Ormuri; New Babilonian; Late Old Umbrian; North Selkup
One place of articulation: Tumshuqese; Old Greek Pamphylian; Parochi, Ormuri

A.2 *IVV*

Three places of articulation: Sarmatian, Alanian, Common Ossetian; Chwaresmian; Khotanese and Tumshuqese; Pamir (Yazghulami, Sughni); Bactrian; Yidgha-Munji ; Pashto; Common North-West Middle Iranian; Middle Persian, Common South-West Iranian; Karelian (Proper, Olonets, Ludic), Dialectal Ingrian, Veps; Dialectal South Estonian, Livonian; Common North-West Middle Indo-Aryan; Common Middle Indo-Aryan; Central Italian Marche, Umbrian, Latio, New Corsican; Common West Romance; Middle Selkup; Proto-Brittonic; Manx; South Norwegian, Swedisch (Bohuslän, West Västergötland); Common Old Danish; Common Permic; Kamassian, Koibalian; Pamir Sanglechi-Ishkashimi; Gorani, Azari, Sivendi ; Common Nuristani; Mordvinian; Mari; Dialectal South Selkup
Two places of articulation: South-West Iranian Kumzari; North Khanty (Obdorsk); Sogdian; Common Parachi-Ormuri; Enets; Tundra Nenets; South-West Iranian Bashkardi
One place of articulation: Waxi; Common Parachi-Ormuri; Proto-Anatolian

A.3 *FD*

Three places of articulation: Dardic (Kashmiri); Dardic (Dameli); Nuristani (Askunu, Kati, Prasun); Common Rhaeto-Romance, Common French, Franco-Provençal, Common Occitan,; Breton; Proto-Nordic; Old Low Franconian, Old High German Middle Franconian, North Thuringian; New Dutch; Middle Low German, New Low German; Bulgarian; Serbo-Croatian West Chakavian (Istrian), Common Slovenian; Common Slovak; Russian; West Russian; Ethiopian Tigrinya; Dardic (Kashmiri); Common Catalan; Old Bretton; Common Middle High German; New West Frisian
Two places of articulation: Sangsari
One place of articulation:Yaghnobi; Old Saxon; Old High German South Rhine Franconian, East Franconian

A.4 *PND*

A.4.1 *First change*

Three places of articulation: Milian, Lycian, Karian; Common East Middle Iranian; Old Greek (Pamphylian, Central Krete, West Argolian, Laconia, Elis, Boiotic); Late Old Greek Helenic; Common South Italian; Milian, Lycian; Young Avestan; Parachi; Late Old Persian, Common West Iranian; Late Parthian, Common North-West-Iranian; Late Middle Persian, South-West Iranian (Dialectal); Common Celtic; Common Finnic; Mari; Common Akkadian (Dialectal); Middle Hebrew, Phoenician, Aramaic; East Balochi; Central Italian (Tuscany Dialectal); Common Middle Indo-Aryan; New West Middle Indo-Aryan; Common Nuristani; Khotanese and Tumshuqese; Common Parachi-Ormuri; Common New Italian, Common Rhaeto-Romance; Common French, Franco-Provençal; Common Catalan, Common Spanish; Portuguese European; Manx; Old Danish Common ; East Saami; Mordvinian; Permic
Two places of articulation: Central Kurdish (War.); Pamir Sanglechi-Ishkashimi; Ormuri; Yidgha-Munji ; Ethiopic Amharic; New High German , West Alemannic, Nord, East Middle Bavarian; Common North-West Midde Indo-Aryan; Common Middle Indo-Aryan; Dalmatian; Common Sardinian; Common West Romance

One place of articulation: Tigre, Tigrinya, Harari, Gurage; West Tocharian Dialectal; Common Middle Indo-Aryan; Ossetian; Late Baktrian; Old Greek Pamphylian; Vulgar Latin, Common Rommance; New High German Common Middle Bavarian; Central Italian (Tuscany Dialectal); Common Old Swedish; Middle Dutch; Low German (dialectal); New North Frisian (Sylt, Festland); Enets (karassinisch); New High German , South and East Moselle Franconian, Hessian, Palatine; Common Albanian; Pashto; Central Kurdish (Sul.); Umbrian; Yidgha, Munji (dialectal); Central Italian (Tuscany Dialectal); New North Frisian (Dialectal); New Dutch

A.4.2 Second Change

Three places of articulation: New Persian, North-West Iranian Dialectal; Scottish Gaelic, Manx; Common New Icelandic; Common New Danish (dialectal); Old High German East Franconian, Upper German; Proto-Armenian; Proto-Phrygian; Proto-Germanic; Hittite, Palaic, Luwian; Old Indo-Aryan dialectal; New High German Dialectal; Common Anatolian; Common Tocharian
Two places of articulation: South Italian (South Latio, North Camoanian)
One place of articulation: Old High German East Franconian, Upper German; Old High German Rhine Franconian; South Italian (East Apulia, East Sicilian dialectal)

A.4.3 Third Change

Three places of articulation: Slovenian Upper Carniolan (Dialectal)
Two places of articulation: New Persian (Mod.); Italian New Greek (Terra d'Otranto, Bova); Common Rhaeto-Romance; Low German (Ostfalian, Central Low Saxon); Common Old High German
One place of articulation: West New Aramaic; New Low German Westphalian dialectal; New High German South Bavarian West Cimbrian; North-East New Aramaic (Jewish Azerbaijan); Dialectal Arabic Bahrain Shia; Ugaritic; Common South-West Iranian, Old Persian; Pamir Ishkashimi; Yaghnobi; Waxi; Middle Norwegian; Middle Swedish; New Swedish; Common West Germanic; New English (dialectal); New North Frisian (Heligoland); Middle Low German, Middle Dutch; East Saami (Ter); South Saami (southern dialects); Veps, Livonian; Morvinian; South New Irish dialectal (Munster); Ormuri (Baraki Barak); Pashto dialectal (North-East, North-West); Common Old Armenian; New Greek (Bova); Occitan dialectal; New Low German (Mecklenburg); Arabic dialectal (Egyptian, Levantine, Maghrebi, Maltese); New South Arabic (Soqotri); Common Aramaic; North-East New Aramaic; Ossetian; Dialectal Albanian (östl); Late Old Frisian dialectal, Common New West Frisian; New East Frisian (Saterland); New East Frisian (Wangerooge); Lombardic

A.5 IVD

A.5.1 First Change

Three places of articulation: Milian, Lycian; Young Avestan; Parachi; Late Old Persian, Common West Iranian; Late Parthian, Common North-West-Iranian; Late Middle Persian, South-West Iranian (Dialectal); Common Celtic; Common Finnic; Mari; Common Akkadian (Dialectal); Middle Hebrew, Phoenician, Aramaic; East Balochi; Central Italian (Tuscany Dialectal); Common Middle Indo-Aryan; New West Middle Indo-Aryan; Common Nuristani; Khotanese and Tumshuqese; Common Parachi-Ormuri; Central Kurdish (War.); Common New Italian, Common Rhaeto-Romance; Common French, Franco-Provençal; Common Catalan, Common Spanish; Portuguese European; Manx; Old Danish Common ; East Saami; Mordvinian; Permic
Two places of articulation: Pamir Sanglechi-Ishkashimi; Ormuri; Yidgha-Munji ; Ethiopic Amharic; New High German , West Alemannic, Nord, East Middle Bavarian; Common North-West Middle Indo-Aryan; Common Middle Indo-Aryan; Dalmatian; Common Sardinian; Common West Romance;
One place of articulation: Tigre, Tigrinya, Harari, Gurage; West Tocharian Dialectal; Common Middle Indo-Aryan; Ossetian; Late Baktrian; Old Greek Pamphylian; Vulgar Latin, Common Rommance; New High German Common Middle Bavarian; Central Italian (Tuscany Dialectal); Common Old Swedish; Middle Dutch; Low German (dialectal); New North Frisian (Sylt, Festland); Enets (karassinisch); New High German , South and East Moselle Franconian, Hessian, Palatine; Common Albanian; Pashto; Central Kurdish (Sul.);

Umbrian; Yidgha, Munji (dialectal); Central Italian (Tuscany dialectal); Dialectal Selkup; New North Frisian (dialectal)

A.5.2 Second Change

Two places of articulation: New Dutch (dialectal); Spanish Aragonese Castillian, Andalusian; New Danish (South Jutland); Palatine, East Franconian, Upper Saxon dialectal
One place of articulation: Galician; Catalan (Apixtat), Ribagorza; Young Avestan, West Iranian (dialectal); New High German (Middle Franconian); New Low German dialectal; East and South Italian (dialectal)

A.5.3 Third Change

Three places of articulation: Waxi; Balochi
Two places of articulation: Central Italian (Tuskany dialectal)
One place of articulation: North-East New Aramaic; Arabic dialectal (Egyptian, Levantine, Maghrebi, Maltese); New South Arabic (Soqotri); Common Aramaic; North-East New Aramaic; Ossetian; Dialectal Albanian (östl); Late Old Frisian dialectal, Common New West Frisian; New East Frisian (Saterland); New East Frisian (Wangerooge); Lombardic; Pamir Sanglechi; New Greek Anatolian; Central Sardinian, Logudorese; Middle Norwegian, Middle Swedisch, Common Old Danish; New English (Shetland, Orkney, Manx); South Khanty, North Khanty (Nizjam, Sherkaly); Mansi; Serbo-Croatian Montenegrin; New Greek dialectal (Bova); Parochi

A.6 FV

A.6.1 First Change

Three places of articulation: Hebrew; Amharic; Old Saxon, Old Dutch, Old High German

A.6.2 Second Change

Three places of articulation: Sarmatian, Alanian, Common Ossetian; Chwaresmian; Khotanese and Tumshuqese; Bactrian; Yidgha-Munji ; Pashto; Common North-West Middle Iranian; Middle Persian, Common South-West Iranian; Karelian (Proper, Olonets, Ludic), Dialectal Ingrian, Veps; Dialectal South Estonian, Livonian; Middle Selkup; South Norwegian, Swedisch (Bohuslän, West Västergötland); Common Old Danish; Kamassian, Koibalian; Gorani, Azari, Sivendi ; Mordvinian; Mari; Dialectal South Selkup
Two places of articulation: South-West Iranian Kumzari; Sogdian; Common Parachi-Ormuri; Enets; South-West Iranian Bashkardi; Tundra Nenets
One place of articulation: Common Parachi-Ormuri; Proto-Anatolian

A.6.3 Third Change

Three places of articulation: Ethiopian Gurage; Milian, Lycian; Common Dardian; Sindhi; New Greek; Common Albanian; Romanian; Common North Italian; Common Rhaeto-Romance, Common Gallo-Romance; Common Ibero-Romance; Old Irish; Common Brittonic ; Danish Jutlandic, Insular; Middle English; Common New Frisian, Middle Low German, Middle Dutch ; Mordvinian, Common Mari; Common Permic; Common Khanty, Mansi; Hungarian; Proto-Samoyedic; Common New High German

B Synchronic typology

B.1 Post-nasal voicing

Counts are based on a query for a change from [−voiced] to [+voiced] with the left environment conditioned on [+nasal] in the P-base database in Mielke (2019).

Mixe, Lowland (Coatlán variety); Maasai; Binumarien; Delaware (Unami); Pitjantjatjara/Western Desert Language; Guatuso (Maléku Jaíka); Bemba; Khmu?; Xhosa; Totonac, Misantla; Mixe, North Highland (Totontepec Mixe); Kui; Quichua, Ecuador (Puyo Pongo variety); Mwera; Passamaquoddy-Maliseet (Malecite-Passamaquoddy); Tamil; Muruwari; Malayalam; Ciyao (Yao); Nyangumata; Si-Luyana; Kpelle; Tiriyó (Trió); Pero (Gwandum dialect); Quechua, Ecuadorean Highland (as spoken in Bolivar Province); Limbu; Ojibwa, Eastern; Pero

B.2 Intervocalic voicing

Counts are based on a query for a change from [−voiced] to [+voiced] with the left environment conditioned on [+vowel] in the P-base database in Mielke (2019).

Af Tunni Somali (Tunni); Alabama; Ao; Auyana; Berbice Dutch Creole; Boruca; Bribri; Burmese; Danish; Efik; Estonian; Faroese; Faroese (in some districts); Guatuso (Maléku Jaíka); Kalenjin, Nandi; Kui; Kwamera; Lele; Loniu; Mangap-Mbula; Martuthunira; Mikasuki; Mixe, Lowland (Coatlán variety); Mixe, Lowland (Guichicovi variety); Mixe, Lowland (San Juan el Paraíso variety); Mixe, North Highland (Totontepec Mixe); Mixe, South Highland (Mixistlán variety); Mixe, South Highland (Tepantlali variety); Mixe, South Highland (Tepuxtepec variety); Mixe, South Highland (Tlahuitoltepec variety); Mohawk; Mupun; Mupun (Jipari dialect); Ngura; Nyangumata; Ojibwa, Eastern; Oneida; Palauan; Passamaquoddy-Maliseet (Malecite-Passamaquoddy); Pech (Paya); Pero; Popoluca, Sayula; Purik; Quechua, Ecuadorean Highland (as spoken in Bolivar Province); Senoufo, Supyire; So (Soo); Tangkhul; Tsimshian, Coast; Turkish; Tyvan (Tuvin); Xakas (Khakas)

B.3 Final devoicing

Counts are based on a query for a change from [+voiced] to [−voiced] with the right environment conditioned on word boundary in the P-base database in Mielke (2019).

Fe'Fe'-Bamileke; Ejagham; Slovene; Shilluk; Czech; Lithuanian; O'odham (Papago); Polish; Pero; Afrikaans; Kirghiz; Tigre; Turkish; Russian; Af Tunni Somali (Tunni); Amele; Ingessana; Boruca; Dutch; Armenian, Standard Eastern; Dhaasanac (Daasanach); Bulgarian; Arbore; Tirmaga; Serbo-Croatian (Cres Čakavian); Wolof; Faroese; Sepečides-Romani; Slovak; Nigerian English (Nigerian Pidgin); Afar;

C Supplementary materials

C.1 bsc()

The function *bsc()* takes two vectors of equal length as arguments: a vector with counts of languages with a sound changes required for an alternation A_k, and a vector of languages surveyed for each sound change. The function internally transforms the vectors with counts into a binomial distribution of successes and failures for each sound change in the count. It returns R bootstrap replicates of the Historical Probability of A_1, computed according to (2), (3), (4), and (5). Stratified non-parametric bootstrapping is performed based on the *boot* package: the output of *bsc()* is an object of class "boot". The output of *bsc()* should be used as an argument of *summary.bsc()* (see C.3), which returns the observed P_χ and 95% BC_a CIs. Two optional arguments of *bsc()* are *order* (if True, Historical Probabilities are divided by $n!$) and R, which determines the number of bootstrap replicates.

```
bsc <- function (counts, surveyed, order = T, R = 10000) {
  library(boot)
  if (length(counts) != length(surveyed)) {stop
    ("Vectors must be of equal length.")
  }
  binom <- unlist(mapply(c,
                         lapply(counts, function(x) rep(1, x)),
                         lapply(surveyed - counts, function(x) rep(0, x)),SIMPLIFY
                           =F)
```

```
  )
  snumb <- paste("s", 1:length(surveyed), sep="")
  ident <- rep(snumb, surveyed)

  scsample <- data.frame(binom,ident)

  if (order == TRUE) {n <- factorial(length(counts))}
  if (order == FALSE) {n <- 1}

  bsc <- function(x, id) {
    sc1 <- tapply(x[id,1], x[id,2], mean)
    sc <- prod(sc1) / n
    return(sc)
  }

  boot.scsample <- boot(scsample, statistic = bsc, R, strata = scsample[, 2]
                        )
  return(boot.scsample)
}
```

C.2 bsc2()

The function *bsc2()* compares the Historical Probabilities of two processes with BSC. It takes as an input the output of *bsc()* for the process in question. The function transforms the counts into a binomial distribution of successes and failures. It returns R bootstrap replicates of the difference in Historical Probability between the two alternations, computed according to (2), (3), (4), (5), and (6). Stratified non-parametric bootstrapping is performed based on the *boot* package: the output of *bsc2()* is an object of class "boot". The output of *bsc2()* should be used as an argument of *summary.bsc2()* (see C.4), which returns the observed ΔP_χ and 95% BC_a CIs for the difference. If 95% BC_a CIs fall above or below zero, it spells out that the difference is significant, and that it is not otherwise. Two optional arguments of *bsc()* are *order* (if True, Historical Probabilities are divided by $n!$) and R, which determines the number of bootstrap replicates.

```
bsc2 <- function(bsc.alt1a, bsc.alt2a, order = T, R = 10000){
  library(boot)
  bsc.alt1 <- bsc.alt1a$data
  bsc.alt2 <- bsc.alt2a$data
  bsc.alt1$scid <- "first"
  bsc.alt2$scid <- "second"
  bsc.diff.df <- rbind(bsc.alt1,bsc.alt2)
  bsc.diff.df$comb <- as.factor(paste(bsc.diff.df$scid,bsc.diff.df$ident, sep = ""
     ))

  bsc.diff.df$scid <- NULL
  bsc.diff.df$ident <- NULL

  if (order == TRUE) {  n1 <- factorial(length(unique(bsc.alt1$ident)))
  n2 <- factorial(length(unique(bsc.alt2$ident)))}
  if (order == FALSE) { n1 <- 1
  n2 <- 1}

  l <- length(unique(bsc.alt1$ident))
  m <- length(unique(bsc.alt2$ident))

  bsc.diff <- function(x, id) {
    sc1 <- tapply(x[id,1], x[id,2], mean)
    sca <- (prod(sc1[1:l]) / n1)
    scb <- (prod(sc1[(l+1):(l+m)]) / n2)
    sc <- sca - scb
```

```
    return(sc)
  }

  boot.diff <- boot(bsc.diff.df, statistic = bsc.diff, R, strata = bsc.diff.df[, 2]
                    )
  return(boot.diff)
}
```

C.3 *summary.bsc()*

The function *summary.bsc()* computes the 95% BC_a CI for the bootstrap replicates based on the *bsc()* function (see C.1) using the *boot.ci()* function from the *boot* package and returns the observed and estimated Historical Probabilities. For details, see C.1.

```
summary.bsc <- function (bsc.alt) {
  bsc.ci.alt <- boot.ci(bsc.alt, type="bca")
  title <- "BOOTSTRAPPING␣SOUND␣CHANGES"
  prob <- paste("Estimated␣P␣=", round(bsc.alt$t0*100, digits = 5), "%")
  bca <- paste("Estimated␣95␣%␣BCa␣CI␣=␣[", round(bsc.ci.alt$bca[4]*100, digits = 4),"%,",
               round(bsc.ci.alt$bca[5]*100, digits = 4),"%]")
  #rnsc <- paste(pasteR, n.sc.paste, countsp, surveyed, sep = "\n")
  probbca <- paste(prob, bca, sep = "\n")
  cat(title, probbca, sep = "\n\n")
}
```

C.4 *summary.bsc2()*

The function *summary.bsc2()* computes the 95% BC_a CI for the bootstrap replicates based on the *bsc2()* function (see C.2) using the *boot.ci()* function from the *boot* package and returns the observed and estimated differences in Historical Probabilities of two alternations. For details, see C.1.

```
summary.bsc2 <- function (bsc2.alt) {
  bsc2.ci.alt <- boot.ci(bsc2.alt, type="bca")
  title <- "BOOTSTRAPPING␣SOUND␣CHANGES␣-␣COMPARE"
  prob <- paste("Estimated",expression(Delta), "P␣=", round(bsc2.alt$t0*100, digits = 5), "%")
  bca <- paste("Estimated␣95␣%␣BCa␣CI␣=␣[", round(bsc2.ci.alt$bca[4]*100, digits = 4),"%,",
               round(bsc2.ci.alt$bca[5]*100, digits = 4),"%]")
  if (bsc2.ci.alt$bca[4] > 0 & bsc2.ci.alt$bca[5] > 0) {
    sig <- "P(A1)␣is␣significantly␣higher␣than␣P(A2)."
  }
  else if (bsc2.ci.alt$bca[4] < 0 & bsc2.ci.alt$bca[5] < 0) {
    sig <- "P(A1)␣is␣significantly␣lower␣than␣P(A2)."
  } else {
    sig <- "P(A1)␣and␣P(A2)␣are␣not␣significantly␣different."
  }
  probbca <- paste(prob, bca, sep = "\n")
  cat(title, probbca,sig, sep = "\n\n")
}
```