Bootstrapping Sound Changes*

Gašper Beguš

Abstract

This paper presents applications of a technique for estimating influences of the Channel Bias on phonological typology called *Bootstrapping Sound Changes* (BSC). Following Author (2017), we argue that for any synchronic alternation, the BSC technique enables estimation of the probability that the alternation arises based on the number of sound changes it requires and their respective probabilities. This paper develops and illustrates further applications of the proposed model. With the Bootstrapping Sound Changes technique, we can compare Historical Probabilities of attested and unattested alternations and perform inferential statistics on the comparison, predict (un)attestedness in a given sample for any alternation, and derive quantitative outputs for a typological framework that models both Channel Bias and Analytical Bias influences together. The BSC technique also identifies several mismatches in typological predictions of the Analytic and Channel Bias approaches. By comparing these mismatches with the observed typology, the 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 discussion of which factors influence the observed typology. Two major lines of thought emerge in this discussion: the Analytic Bias (AB) and Channel Bias (CB) approach (Moreton 2008).[1] The AB approach argues that the observed typology results primarily from differences in the learnability of phonological processes; the CB approach argues that the inherent directionality of sound changes based on phonetic precursors (articulatory and perceptual) results in typology (for a more detailed discussion, see Ohala 1981, 1983, 1993, Kiparsky 1995, 2006, 2008, Hyman 2001, Blevins 2004, Wilson 2006, Moreton 2008, Hayes et al. 2009, Moreton and Pater 2012a,b, de Lacy and Kingston 2013, Cathcart 2015, Greenwood 2016, i.a.).

Argumentation in favor of or against one or the other approach has been primarily evaluated qualitatively rather than quantitatively (Ohala 1981, 1983, 1993, Kiparsky 1995, 2006, 2008, Blevins 2004). This is especially problematic because empirical evidence often supports both lines of thought equally well. Attempts to quantify the influence of one or the other approach on typology have nevertheless been made on both sides of the typological discussion. Most of the proposed models, however, face crucial challenges or fail to yield results that would allow for a quantitative comparison of the two approaches. Below, we briefly outline some challenges of the current proposals before implementing a new quantitative model of phonological typology within the Channel Bias approach.

*I would like to thank Kevin Ryan, Jay Jasanoff, Adam Albright, Edward Flemming, Patrick Mair, Morgan Sonderegger, and the audiences at NELS 48, AMP 2016, 2017, WCCFL 35, and CLS 53 for useful comments on earlier versions of this paper. All remaining mistakes are my own. The research is partially funded by Mind Brain Behavior Interfaculty Initiative at Harvard University.

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

The AB approach argues that learning biases influence typology, and substantiates this claim with evidence from artificial grammar learning experiments. 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. To derive typology from learnability differences is, however, not a trivial task. Using computational modeling, Rafferty et al. (2011) suggest that learnability differences might not be sufficient for deriving typological observations. On the other hand, recent models proposed by Morley (2015) and Stanton (2016) have derived a particular subtype of typological observations from learnability differences. Perhaps the most powerful model so far is the Maximum entropy model of phonological learning (Goldwater and Johnson 2003, Wilson 2006) that computationally encodes learnability differences between different phonological processes. Staubs (2014) proposes a model that derives the observed typology from learnability differences. The ability to derive typological observations computationally does not, of course, constitute sufficient evidence that AB directly influences the observed phonological typology, but it does begin to address the objection raised by Rafferty et al. (2011).

A broader challenge that the AB approach faces, however, is that experiments testing the learnability of typologically rare or nonexistent unnatural processes consistently fail to show learnability differences compared to typologically frequent natural processes when the structural complexity of the tested alternation is controlled for. Influences of the 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 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 at best. Many studies that test unnatural alternations as defined in Section 2 (as opposed to unmotivated alternations) found no effect of Substantive Bias (Pycha et al. 2003, Wilson 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; cf. Carpenter 2006, 2010, Wilson 2006). Deriving substantial typological differences between natural and unnatural alternations within the AB approach is problematic if no differences in learning are observed experimentally.

Attempts to quantify the Channel Bias influences on phonological typology are even more sparse. One of the first attempts in Bell (1970, 1971) and Greenberg (1978) is underdeveloped. The authors propose a "state-process model" for calculating probabilities of transitions from different diachronic states, but estimates of individual transition probabilities are mostly based on informed guesses rather than on typological surveys. Moreton (2008) proposes a model for quantifying phonetic precursors. As will be shown below, however, phonetic precursors and results of sound changes do not always align, which makes the model less appropriate for deriving unnatural alternations. Additionally, quantifying phonetic precursors, as proposed in Moreton (2008), is not a trivial task (for objections against the proposal, see Yu 2011). Our model overcomes this difficulty by estimating probabilities of individual sound changes directly from typological surveys. Finally, Cathcart (2015) proposes a framework that does model combinations of sound changes and operates with typological surveys. The model is, however, computationally too demanding to yield implementable results. One problem with the model is that it crucially relies on the representativeness of surveys of all possible sound changes, not only of the ones required for an alternation in question. As will be

[2]Complexity Bias has also been called Substantive Bias.

shown below, our model requires surveys of sound changes to be representative only for the sound changes required for a given alternation. The model in Cathcart (2015) also fails to distinguish alternations from static phonotactic requirements.

The goal of this paper is to propose a quantitative method for estimating influences of Channel Bias on phonological typology. The model proposed here is based on the concepts Minimal Sound Change Requirement (MSCR), Historical Probabilities of Alternations (P_χ), and Bootstrapping Sound Changes, developed in Author (2017). This paper proposes an elaborated version of the model and illustrates several of its new applications. The paper argues that with the BSC 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), (iii) predict attestedness in a given sample for any alternation (Section 4.3), and (iv) derive a quantitative metric for the Channel Bias influences on typology that can be employed in typological frameworks that model both AB and CB together (Section 4.4). Using the BSC technique, we also identify and quantify several predictions of the Channel Bias approach and compare them with predictions of the Analytic Bias approach (Sections 4.4.2 and 5). This allows us to identify crucial mismatches in predictions of the two approaches which in turn allows for disambiguation between AB and CB influences on observed typology (Section 5).

The paper applies the BSC technique to three unnatural alternations that target the feature [±voice]: post-nasal devoicing (PND), intervocalic devoicing (IVD), and final voicing (FV). The feature [±voice] is chosen for several reasons. First, phonetic naturalness is probably best understood precisely for this feature. [+voice] is natural intervocalically and post-nasally and unnatural word-initially and word-finally for clear articulatory and perceptual reasons (Aerodynamic Voicing Constraint; Ohala 1983, 2011, Westbury and Keating 1986, Author 2017). Second, all three alternations are well-researched typologically: PND and FV are probably two of the most widely discussed alternations in the phonological literature (Hyman 2001, Kiparsky 2006, 2008, Blevins 2004, Yu 2004, Coetzee and Pretorius 2010). Third, the three alternations crucially differ in their synchronic attestedness: PND is attested as a productive synchronic alternation (Coetzee and Pretorius 2010), IVD is attested as a gradient phonotactic restriction (Author and Name 2017), and FV is arguably never attested as a synchronic alternation (Kiparsky 2006, 2008, cf. Haspelmath 1993, Yu 2004). Fourth, the natural counterparts of the three unnatural alternations are recurrent and typologically common phonetic tendencies and alternations. Finally, two of the three alternations arise from a combination of three natural sound changes (the Blurring Process; see Sections 2 and 4.1) and this development is historically directly or indirectly attested. For FV, Kiparsky (2006) identifies several diachronic trajectories that would yield the alternation, but none appears to be attested. These different degrees of synchronic and diachronic attestedness between PND, IVD, and FV and their respective natural counterparts allow for a comparison of different approaches to phonological typology.

2 Background

Several key concepts from Author (2017) are adopted in this paper. This section outlines these concepts and discusses how they are relevant to the BSC technique. In the interest of space, not all details can be discussed here; the reader should be directed to Author (2017) for further discussion on the Blurring Process and Minimal Sound Change Requirement.

This paper first adopts the division of phonological processes into *natural*, *unmotivated*, and *unnatural*. 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. Universal phonetic tendencies are defined as phonetic tendencies that have articulatory or perceptual motivation, operate passively cross-linguistically, and result in typologically common phonological processes (Author 2017). Natural processes, such as final devoicing or post-nasal and inter-vocalic voicing, are cases of phonetically well-motivated universal phonetic tendencies. Examples of unnatural alternations include final voicing or post-nasal and intervocalic devoicing (for articulatory and perceptual argumentation on the unnaturalness of these processes, see Westbury and Keating 1986, Ohala 1983, 2011). An example of an unmotivated process would be Eastern Ojibwe "palatalization" of /n/ to [ʃ] before front vowels (Buckley 2000).

Sound change is defined as a non-analogical "change of one [non-redundant (non-automatic)] feature in a given environment" (Author 2017). Redundant features that change together with the non-automatic feature do not count as instances of sound change (e.g. a change in [±nasal] automatically causes a change in [±sonorant]).[3] Sound change is also defined as a completed event that ideally targets all vocabulary items in a given language L. Because the BSC technique models sound changes that via phonologization result in phonological alternations, we adopt the level of abstraction from phonology where features encode non-automatic, language-specific, and speaker-controlled alternations that cannot be attributed to universal phonetics (Hyman 2013). It is assumed that a single sound change can only change one feature value (or delete, insert, or reorder a whole feature matrix in case of deletion, epenthesis, or metathesis) in a given environment. This "minimality principle", first proposed in Donegan and Stampe (1979) and Picard (1994), is discussed at length in Author (2017). We also define a *combination of sound changes* as a set of such individual sound changes that each target a single feature value.

The paper also adopts two key concepts in the derivation of typology within the Channel Bias approach that have been proposed in Author (2017): the Minimal Sound Change Requirement and Historical Probabilities of alternations (P_χ). 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, and two additional cases of unnatural phonotactic restrictions (the distribution of the feature [±voice] in Tarma Quechua and in the Berawan dialects) (summarized in Table 1). Author (2017) argues that PND arises from a combination of three sound changes in all reported cases (as was proposed for Tswana in Dickens 1984 and Hyman 2001) and argues that the three-sound-changes approach is historically directly confirmed by Avestan, Sogdian and Yaghnobi (three languages in ancestral relationship that have all three sound changes attested in written sources). Based on this typological survey, a new strategy for diachronically explaining unnatural processes is proposed: the Blurring Process. The Blurring Process states that unnatural alternations arise through a combination of a specific set of three 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. All reported cases of unnatural phonotactic restrictions (in the Berawan dialects and Tarma Quechua) also arise through a combination of three sound changes — again, the Blurring Process.

The typological survey of diachronic developments of unnatural processes together with the newly established Blurring Process reveal another crucial aspect of unnatural alternations: the so-called Minimal Sound Change Requirement. A minimum of three sound changes are required for an unnatural process to arise and a minimum of two sound changes for an unmotivated process (for a formal proof of the MSCR, see Author 2017). As will be shown below, the MSCR has broad implications for deriving typology within the Channel Bias approach.

[3] Such automatic changes can also be language-specific.

Table 1: Unnatural processes and a selection of languages in which they appear.

Unnatural process or sound change	**Language**	**Description**
Post-nasal devoicing (PND)	Yaghnobi	Xromov (1972)
	Tswana and Shekgalagari	Solé et al. (2010)
	Makhuwa and Bube	Janson (1991/1992)
		Janssens (1993)
	Konyagi	Merrill (2016a,b)
	Sicilian and Calabrian	Rohlfs (1949)
	Murik, Buginese	Blust (2013)
	Nasioi	Brown (2017)
Intervocalic devoicing (IVD)	Berawan and Kiput	Blust (2005, 2013)
Post-obstruent stop voicing	Tarma Quechua	Adelaar (1977)

Finally, we adopt the concept of Historical Probabilities of alternations (P_χ) which states that for every synchronic alternation we can estimate its Historical Probability, i.e., the probability that the alternation arises based on the number of sound changes required for that alternation (MSCR) and their respective probabilities.

(1) *Historical Probabilities of alternations* (P_χ)
"The probability that an alternation arises based on the number of sound changes required (MSCR) and their respective probabilities that can be estimated from samples of sound changes."

Author (2017) outlines a technique for estimating Historical Probabilities called Bootstrapping Sound Changes, but illustrates only one of its applications. The section below presents an elaboration of the BSC technique that introduces some key new concepts and presents several novel applications of the BSC.

3 Implementation

3.1 Sample

Samples used for estimating Historical Probabilities with BSC are created from typological surveys of sound changes. The BSC 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 occurrence 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 based on which we perform the BSC analysis 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 well-reconstructed prehistory and well-established subgrouping. This allows the survey to properly code the occurrence of a sound change, where 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 or 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 sound change as well as from higher level evidence (e.g. morphology). Additionally, phylogenetic tree analysis does not restrict the direction of sound change and would crucially analyze unnatural alternations as resulting from a single sound change.

The survey in Kümmel (2007) includes approximately 294 languages and dialects of Indo-European, Semitic, and Uralic language families. While the survey is not as representative because it excludes a large number of 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 our analysis 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 (see also Section 4.4.2).

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 for the BSC technique 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. In addition to the lack of subgrouping, the UniDia database is not representative either, focusing primarily on the Bantu language family (83.5% of sound changes are from the Bantu family).

The BSC technique offers some crucial advantages over other quantitative approaches to estimating probabilities of sound changes or their combinations, especially over the one proposed in Cathcart (2015). The requirement for samples to be representative is much weaker under the BSC approach. Cathcart's (2015) model crucially requires surveys of sound changes to be representative for all possible sound changes. Because identification of historical trajectories that lead to an alternation is performed manually in our model, surveys of sound changes that we use for BSC 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. Additionally, Cathcart (2015) uses the UniDia database for his model, which is less appropriate compared to Kümmel's (2007) survey, primarily because of its encoding of sound changes that lacks subgrouping.

3.2 Bootstrapping

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. It was first proposed in Efron (1979) and has seen a wide range of applications ever since (Davison and Hinkley 1997).

The model uses a stratified non-parametric bootstrap technique for estimating Historical Probabilities for several reasons. First, the statistic of interest in BSC is often too complex for an easy analytic solution, especially when we estimate Historical Probabilities of alternations that require more than a single sound change (4) or when we estimate differences between two Historical Probabilities (see 4.2 below). Second, bootstrapping is a frequentist technique for estimating sampling distribution for a statistic of interest and as such requires no prior beliefs. Finally, bootstrapping allows for inferential statements on the comparison of Historical Probabilities of two alternations, even when the statistic of interest is complex (as will be shown below).

The computation of BSC is implemented in the R Statistical Software (R Core Team 2016) with the *boot* package (Canty and Ripley 2016, Davison and Hinkley 1997) using functions *boot()* and *boot.ci()*. This paper also presents R code that implements the BSC technique and introduces functions *bsc()*, *summary.bsc()*, *bsc2()*, *summary.bsc2()*, *plot.bsc()* and *plot.bsc2()* (based on the *boot* package) that facilitate the estimation of Historical Probabilities with BSC (available in A). The functions allow estimation of Historical Probabilities directly from a vector of counts and should be easy to use even for researchers without substantial statistical knowledge. The aim of the code is to provide an interface for the estimation of the Historical Probability of any alternation and thus to provide a means for estimating the Channel Bias influence in future discussions on phonological typology.

3.3 The model

As defined in (1), the Historical Probability of an alternation A_k is the probability that a language L features A_k based on the number of sound changes (S_i) the alternation A_k requires and their respective probabilities.

3.3.1 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 (i.e. A_k is natural), then we estimate its P_χ according to (2).

(2)

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

The BSC samples with replacement from the sample of successes and failures and calculates the statistic of interest: in our case, the probability according to (2). This is repeated 10,000 times, which yields a sampling distribution of Historical Probabilities: 10,000 data points. From this sampling distribution we compute standard error, bias, and 95% adjusted bootstrap (BC_a) confidence intervals that adjust for bias and skewness (Efron 1987).

The analytic equivalent of the BSC technique for an alternation that requires only a single sound change is an empty logistic regression model with the number of successes and failures as the dependent variable and with only the intercept with no predictors. As the statistic of interest becomes more complex when estimating Historical Probabilities of processes that require multiple sound changes, we shift from the analytic framework to a non-parametric bootstrap. For consistency, we maintain the BSC approach even for alternations that require only a single sound change and could otherwise be estimated using an analytic approach.

3.3.2 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; see (3).

(3)

$$P_\chi(A_k) = P_\chi(T_1 \cup T_2 \cup T_3 \cup \ldots \cup T_n)$$

A trajectory T_j denotes a combination of sound changes that yields an alternation A_k. In theory, there is 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 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 a 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; see (4).

(4)

$$P_\chi(T_j) = \frac{P_\chi(S_1 \cap S_2 \cap S_3 \cap \ldots \cap S_n)}{n!}$$

Estimating the joint probability of individual sound changes ($P_\chi(S_1 \cap S_2 \cap S_3 \cap \ldots \cap S_n)$) is not a trivial task. We need to make a number of assumption 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, we treat sound changes as independent events under the BSC. For a discussion of assumptions of the BSC model, see Section 3.4 below.

As defined in (1), 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 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 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 the probability that the alternation A_k arises in a language L (regardless of whether it features stop geminates) and *not* the probability that the alternation A_k arises in a language L that features geminate stops.

Once S_1 operates, however, we know that language L necessarily features the target/result/-context of the sound change S_1. For this reason, we estimate the Historical Probability of the subsequent sound changes $P_\chi(S_2)$ from the number of successes (languages with S_2) divided 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 feature the target/result/context of the previous sound change, we can treat the probabilities of individual sound changes as independent events and estimate P_χ from a product of probabilities of individual sound changes:

(5)

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

To estimate standard errors or BC_a confidence intervals for a Historical Probability of A_k that requires more than a single sound change, the BSC technique samples with replacement from n number of 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 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, based on which standard errors and BC_a confidence intervals are computed.

3.3.3 Comparison

BSC not only estimates Historical Probabilities of individual alternations, but also allows for the estimation of the difference between the Historical Probabilities of two alternations.

(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 3.3.1 and 3.3.2 (depending on whether A_1 and A_2 require trajectories that require one or more sound changes). Then using BSC on the difference additionally calculates the difference between $\text{P}_\chi(\text{A}_1)$ and $\text{P}_\chi(\text{A}_2)$, which returns 10,000 bootstrap replicates, based on which standard errors and BC_a confidence intervals are computed.

The BSC technique applied on a difference between two alternations enables 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 $\text{P}_\chi(\text{A}_1)$ and $\text{P}_\chi(\text{A}_2)$ are significantly different with $\alpha = 0.05$. If, on the other hand, the 95% BC_a confidence intervals of the difference cross 0, then $\text{P}_\chi(\text{A}_1)$ and $\text{P}_\chi(\text{A}_2)$ are *not* significantly different with $\alpha = 0.05$.

3.4 Assumptions

The model presented in Section 3.3 makes some crucial assumptions that are discussed in this section. In order to estimate the joint probability of two or more sound changes as a product of Historical Probabilities (see (5)), the model assumes that each sound change is an independent event. This is not a controversial assumption: there is no reason to believe occurrence of one sound change affects the probability of the following sound changes, unless the first sound change crucially alters phonemic inventory of the language in question. The BSC model does, however, account for at least some of the dependency between sound changes and phonemic inventories by estimating probabilities from samples conditioned on the result of previous sound change and by evaluating sound changes according to the target, result, and context.

What is not accounted for in the model are the functional load of individual phonemes and the dependency of sound changes on broader phonemic inventories that do not immediately affect the target, result, or context of the sound changes in question. Broader phonemic inventories can influence probabilities of sound changes, especially for vocalic changes (due to the effects described in the Theory of Adaptive Dispersion, see Liljencrants and Lindblom 1972, Lindblom 1990). BSC also does not model other factors that could potentially influence probabilities of sound changes, such as language contact or sociolinguistic factors, and makes no assumptions about how sound change is initiated or spread.

Finally, the BSC technique does not directly model the temporal dimension. If more comprehensive typological studies with more detailed temporal information were available, a different model (e.g. a model operating within the Poisson stochastic process) could account for the temporal dimension and estimate probabilities of sound changes given a timeframe. In the absence of temporal information, the BSC technique has to make some assumptions. First, the probabilities in BSC are estimated 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 have to operate for the resulting alternation to be productive. For example, 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 in its daughter languages. This is a closer approximation to reality than to assume that 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 which is substantiated by empirical evidence: sound change operates and then ceases to operate, at which point it can occur again (e.g. on novel morphological or loanword material). These assumptions about the temporal dimension are not limited to our model: any model of sound change probability will be faced with these problem because of the lack of more comprehensive surveys that include temporal information. The assumptions in the BSC model approximate the reality reasonably well and, to our knowledge, better than other proposals.

Most of the influences that are not directly modeled in our proposal are at least partially accounted for by the fact that the sample size in our case 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, we can disregard these influences, first because the effects are likely minor enough not to crucially alter our results, and second because current typological surveys do not allow for models that would account for these minor influences. In addition, the BSC technique estimates Historical Probabilities of alternations in a language L, where L represents a language that has the characteristics of the majority of languages in our sample. We do not condition Historical Probability on its phonemic inventory, functional load of phonemes, or other factors, which is why we can disregard these other factors for practical purposes.

4 Applications

4.1 Estimation of Historical Probabilities

The BSC technique enables estimation of Historical Probabilities for any synchronic alternation. We can estimate Historical Probabilities of natural, unmotivated, and unnatural alternations, both attested and unattested (according to Section 3.3). For the purpose of illustrating the method, we estimate the Historical Probabilities (P_χ) of the natural alternations post-nasal voicing (PNV), final devoicing (FD), and intervocalic voicing (IVV), and their unnatural counterparts post-nasal devoicing (PND), final voicing (FV), and intervocalic devoicing (IVD). These processes have received a substantial amount of attention in phonological literature, show different degrees of historical and synchronic attestedness, and are typologically, phonetically, and experimentally well researched (see Section 1).

To construct samples of sound changes for these six alternations, we use the survey of consonantal sound changes in Kümmel (2007). The three natural alternations have the 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 (for a definition, see Section 2) that yield the alternation in question. If A > B / X is a natural sound change, B > A / X is unnatural. (7), (8), and (9) 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 and Author 2017) and identify the actual sound changes that yield the unnatural alternation (right column).

Author (2017) demonstrates that PND results from the Blurring Cycle. A combination of the following three natural and well-motivated sound changes yield PND: fricativization of voiced stops that operates in non-post-nasal position, unconditioned devoicing of voiced stops, and occlusion of voiced fricatives to stops. (7) illustrates the development.[4]

	Blurring Cycle — schematic	*PND*
(7)	B > C / −X	D > Z / [−nas]__
	B > A	D > T
	C > B	Z > D
	B > A / X	D > T / [+nas]__

Author and Name (2017) argue that IVD results from the Blurring Chain. Voiced stops fricativize intervocalically, voiced fricatives devoice, and voiceless fricatives get occluded to stops (see (8)). The result is the unnatural intervocalic devoicing (D > T / V__V).

	Blurring Chain — schematic	*IVD*
(8)	B > C / X	D > Z / V__V
	C > D	Z > S
	D > A	S > T
	B > A / X	D > T / V__V

FV is arguably unattested both as a synchronic alternation as well as a sound change (Kiparsky 2006, cf. Yu 2004). A number of diachronic scenarios, however, exist 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 rather in a static phonotactic restriction. One possible scenario that would result in FV is Scenario 1 in Kiparsky (2006) which we use here for estimating the Historical Probability of FV (Kiparsky 2006).[5] 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 (9)).

	Modified Blurring Cycle — schematic	*FV*
(9)	C > B / X	Tː > T / __#
	B > A	T > D / V__
	C > B	Tː > T
	B > A / X	T > D / __#

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

[5] For a discussion on Scenario 2, see Author (2017).

Based on the trajectories identified here that result in natural and unnatural alternations, we perform counts of sound changes and languages surveyed from Kümmel (2007). PNV that targets labials, dental/alveolars, or velars is reported in approximately 42 languages in Kümmel (2007). IVV is reported in approximately 28 languages if we count only contexts that strictly require intervocalic (as opposed to post-vocalic) context. FD is reported for approximately 33 languages. 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. In fact, sound changes for all six natural and unnatural alternations are counted as successes even if they target only a single place of articulation, because the resulting alternation would count as natural/unnatural, even if it targeted only a single place of articulation. Unclear cases marked with "?" in Kümmel (2007) are excluded from the count. Table 2 summarizes counts of languages with sound changes that result in natural alternations.

Table 2: Counts of sound changes in Kümmel (2007) for natural alternations

Alternation	**Sound change**	**Count**	**Surveyed**
PNV	T > D / N__	42	294
IVV	T > D / V__V	28	294
FD	D > T / __#	33	294

For the unnatural alternations that require more than a single sound change, we perform counts for each individual sound change in the Blurring Process. Fricativization of voiced stops is reported in approximately 97 languages. We include instances of intervocalic and post-vocalic fricativization in the count as well (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.[6] We estimate the probability of the first sound change in the Blurring Cycle that results in PND based on the number of successes (languages in the survey with that sound change) and the total number of language surveyed (294) without conditioning on the sample. The sample for estimating the probability of the first sound changes is unconditioned, because the Historical Probability of A_k is the probability that A_k arises in a language L, regardless of properties of its phonemic inventory (see Section 3.3.2). Once the first sound change operates, however, we know that the language in question needs to feature voiced stops in its inventory. We therefore estimate the Historical Probability of the second sound change that targets voiced stops 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 18 languages (also counting cases of devoicing which are the result of chain shifts). Approximately 31 languages lack voiced stops in the survey in Kümmel (2007),[7] which means that we estimate P_χ based on $294 - 31 = 263$ languages surveyed. After the two sound changes operate, we also know that the language L features voiced fricatives. We estimate P_χ of the last sound change 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,[8] according to Kümmel (2007). In

[6] 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.

[7] One language features only /b/ in its inventory

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

Table 3: Counts of sound changes in Kümmel (2007) for natural alternations

Alternation	Sound change	Count	Surveyed
	D > Z / [−nas]/V__(V)	97	294
PND	D > T	18	263
	Z > D	27	216
	D > Z / V__ (V)	83	294
IVD	Z > S	7	216
	S > T	34	248
	Tː > T / __#	6	294
FV	T > D / V__	32	294
	Tː > T	27	≈88

approximately 27 languages occlusion of fricatives is reported as a sound change.

The first sound change in the Blurring Chain that yields IVD is fricativization of voiced stops post- or intervocalically, which is attested in approximately 83 languages. Because fricativization of voiced stops is the first sound change in the combination, we estimate its probability based on the total number of languages in the survey. The second sound change, unconditioned devoicing of voiced fricatives, is attested in approximately 7 languages out of approximately 216 languages that feature voiced fricatives in their inventories. Finally, occlusion of voiceless fricatives to stops is reported in approximately 34 languages out of 248 languages with non-strident voiceless fricatives in their phonemic inventories.

Counts of the sound changes that lead to FV are the following: in approximately 6 languages, word-final geminates are reported to simplify to singleton stops. 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, we estimate its Historical Probability from the total number of languages surveyed. The second sound change, post-vocalic voicing of voiceless stops, is reported in approximately 32 languages. Because all languages feature voiceless stops, we include all 294 languages surveyed in estimating the Historical Probability of the second sound change. Finally, simplification of geminates is reported in 27 languages. It is difficult to estimate how many languages in Kümmel (2007) allow geminate voiceless stops. While few languages feature phonologically contrastive geminates, many more must allow allophonic geminates at morpheme boundaries. To estimate the number of languages that allow allophonic geminates, we use Greenberg's (1965) survey of consonantal clusters and Ryan's (forthcoming) survey of phonemic geminates. At least 30% of languages in Greenberg's (1965) survey of approximately 100 languages allow stop + stop final clusters. We assume that the number of languages in our sample that allow allophonic homorganic stop-stop sequences (geminates) can be approximated from the proportion of languages that allow sequences of stops or from the proportion of languages that allow phonemic geminates. 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. We thus estimate the number at 88 (30% of 294 languages). That our estimate is accurate is suggested by a survey of phonemic geminates: Ryan (forthcoming) estimates that approximately 35% of 55 genealogically diverse languages surveyed feature phonemic geminates.

To compute the estimates of Historical Probabilities we use the *bsc()* function (see A.1) that transforms two vectors of length n (number of sound changes), where the first vector includes counts of languages in a sample with sound changes in a given trajectory and the second vector includes counts of languages surveyed for each sound change, into a series of successes and failures

Table 4: Estimated P_χ (in %) for natural and unnatural alternations with 95% BC_a and Profile confidence intervals.

		95% $\mathbf{BC}_a$ CI		**95% Profile CI**	
$\mathbf{A}_k$	$\mathbf{P}_\chi$	**Lower**	**Upper**	**Lower**	**Upper**
PNV	14.3	10.2	18.4	10.6	18.6
PND	0.05	0.02	0.09	—	
IVV	9.5	6.1	12.9	6.5	13.2
IVD	0.02	0.008	0.05	—	
FD	11.2	7.8	15.0	8.0	15.2
FV	0.01	0.004	0.03	—	

from which bootstrap replicates are sampled. Based on the *boot* package (Canty and Ripley 2016, Davison and Hinkley 1997), the *bsc()* function performs bootstrapping for the statistic in (2) (if trajectory T_j requires a single sound change) or for the statistic in (5) (if trajectory T_j requires more than one sound change) and returns 10,000 bootstrap replicates. The *summary.bsc()* function computes 95% BC_a confidence intervals based on the bootstrap replicates (using the *boot.ci()* function from the *boot* package; Canty and Ripley 2016, Davison and Hinkley 1997). Table 4 shows the Historical Probabilities with estimated 95% BC_a confidence intervals for the six natural and unnatural alternations discussed above. Figure 1 shows distributions of bootstrap replicates for the Historical Probabilities (P_χ) of these natural and unnatural alternations. Table 4 and Figure 1 illustrate a substantial difference in Historical Probabilities between the natural and unnatural group. The Channel Bias approach estimated with the BSC technique thus predicts that the unnatural alternations (PND, IVD, and FV) will be substantially less frequent than the respective natural alternations (PNV, IVV, and FD).

The Historical Probabilities and confidence intervals of the natural alternations PNV, IVV, and FD could also be estimated analytically (see Section 3.3). To illustrate the accuracy of the BSC technique, we compare the 95% BC_a bootstrap confidence intervals with confidence intervals computed with an analytic solution (see Section 3.3.1). Analytic profile confidence intervals are computed from an empty logistic regression model with a binomial distribution based on the number of successes and failures (languages with and without sound change S_i) and with only an intercept. Table 4 compares the two sets of confidence intervals. The highest difference between the analytic Profile CIs based on a logistic regression model and the BC_a bootstrap CIs is 0.4%, which means the BSC model estimates CIs with high accuracy.

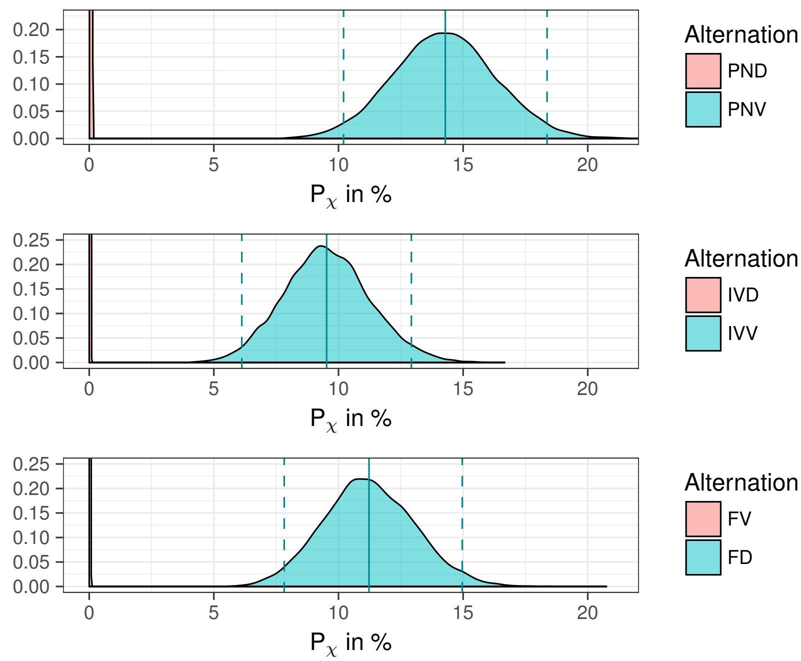

Figure 1: Bootstrap replicates for natural and unnatural alternations. The plots also show the observed P_χ (solid line) and 95% BC_a CI (dashed line). The distribution of bootstrapped P_χ for unnatural alternations does not feature confidence intervals because the probabilities are to small to be visible. For the purpose of representation, the vast majority of bootstrap replicates for unnatural alternations fall outside the limits of the plot.

4.2 Comparison of alternations

One of the advantages of the BSC method is that we can perform inferential statistics on comparisons between the Historical Probabilities of any two alternations. In other words, BSC allows us to test whether the Historical Probabilities of two alternations are significantly different. First, we estimate the difference between the Historical Probabilities of two alternations (ΔP_χ) by stratified non-parametric bootstrap. Samples of sound changes for estimating differences in Historical Probabilities are the same as the samples for estimating Historical Probabilities of individual alternations. BSC samples with replacement and computes the statistic in (6): the difference between two Historical Probabilities. Each individual P_χ is computed as outlined in Section 3.3. If the 95% BC_a confidence intervals for ΔP_χ of two alternations fall either below or above 0, the Historical Probabilities of the two alternations are significantly different (with $\alpha = 0.05$). Functions that perform this computation are *bsc2()* (performs stratified non-parametric bootstrap based on the *boot()* function; see A.2) and *summary.bsc2()* (performs computation of confidence intervals based on the *boot.ci()* function; see A.4).

Figure 2 shows bootstrap replicates of individual Historical Probabilities of 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 with BSC, we can test, for example, whether P_χ(PND) and P_χ(IVD) or P_χ(PND) and P_χ(FV) are significantly different. We estimate the following $\Delta P_\chi(\text{PND}, \text{IVD})$ and $\Delta P_\chi(\text{PND}, \text{FV})$ as described above and in Section 3.3.

(10) a. $\Delta P_\chi(\text{PND}, \text{IVD}) = P_\chi(\text{PND}) - P_\chi(\text{IVD}) = 0.026\%\ [-0.004\%, 0.064\%]$
b. $\Delta P_\chi(\text{PND}, \text{FV}) = P_\chi(\text{PND}) - P_\chi(\text{FV}) = 0.036\%\ [0.011\%, 0.074\%]$

Because the 95% BC_a CIs of the difference in Historical Probability between PND and FV lie above zero, we can claim that P_χ(PND) is significantly higher than P_χ(FV) (with $\alpha = 0.05$). The Historical Probabilities of IVD and PND are, however, not significantly different: the 95% BC_a CIs cross zero. These inferential statements and predictions of the BSC method can be compared

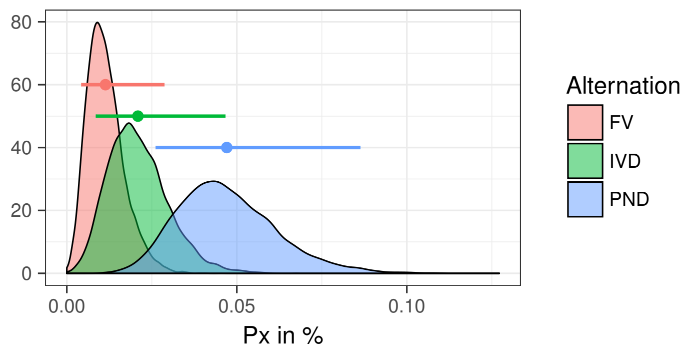

Figure 2: Bootstrap replicates for unnatural alternations with observed P_χ (colored dot) and 95% BC_a confidence intervals (colored lines).

to the observed typology. Kümmel (2007) reports that PND is attested once as a sound change — in Yaghnobi. As was argued in Author (2017), PND in Yaghnobi in fact results from precisely the three sound changes in the Blurring Cycle (7) that we model with BSC. In other words, a combination of sound changes that lead to PND is attested once in Kümmel's (2007) survey. While FV is also reported once in Kümmel (2007), i.e. an alleged *t surfaces as [d] word-finally in Latin, the reported case of FV targets only one series of stops and it is questionable if this sound change was indeed phonetically realized as final voicing (see Kiparsky 2006). Kümmel's (2007) survey also reports a sound change that almost qualifies as IVD: South Italian *d is reported to change to [t] between two unstressed vowels. The scope of this sound change is, however, prosodically limited and its phonetic status is unconfirmed. PND as a diachronic development that results from the Blurring Process seems to be most securely attested in Kümmel's (2007) survey, which is in line with the predictions of the BSC technique.

Looking at the typology more broadly, BSC makes some further desirable predictions. 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 13 languages. In at least two languages the Blurring Process results in a productive synchronic alternation (Table 1). IVD is reported in three languages as a result of the Blurring Process and once as a gradient phonotactic restriction (according to the survey in Author and Name 2017, see Table 1), but never as a productive synchronic alternation. Finally, FV is, to the authors knowledge, never reported as a synchronic phonological alternation (for reasons why Lezgian is not analyzed as featuring FV, see Kiparsky 2006), and only in doubtful cases as a sound change. The typology thus suggests that PND is indeed the most frequent and FV the least frequent, just as predicted by BSC. For evaluation of further typological predictions of the BSC technique, see Sections 4.4.2 and 5.

4.3 Prediction of attestedness

In addition to comparing Historical Probabilities of alternations and performing inference on the comparison, the BSC technique allows for prediction of attestedness in a given sample. We can estimate the difference between the Historical Probability of an alternation A_k and the probability of being attested once in a given sample (with N languages surveyed). If the estimated 95% BC_a confidence intervals fall below or above zero, we can claim that Historical Probability of A_k is significantly higher or lower than the probability of being attested once in a given sample.

Table 5: Estimated ΔP_χ (in %) for natural and unnatural alternations with 95% BC_a confidence intervals.

		95% BC_a CI		
$\mathbf{A}_k$	$\Delta\ \mathbf{P}_\chi$	**Lower**	**Upper**	
PNV	13.9	10.4	18.6	*
PND	−0.3	−1.7	0.05	
IVV	9.2	6.1	13.1	*
IVD	−0.3	−1.8	0.03	
FD	10.9	7.6	15.1	*
FV	−0.3	−2.0	0.01	

(11)

$$\Delta P_\chi(\mathrm{A}_k, \frac{1}{\mathrm{N}}) = P_\chi(\mathrm{A_k}) - P(\frac{1}{\mathrm{N}})$$

The procedure for estimating the difference between the Historical Probability of an alternation A_k and the probability of being attested once in a sample is exactly the same as the procedure for estimating differences in the Historical Probabilities of two alternations (described in 4.2), except that the probability of being attested once in a sample is estimated from a sample of one success and $N-1$ failures. Estimates for each of the six natural/unnatural alternations are computed with functions *bsc2()* and *summary.bsc2()* and are summarized in Table 5. None of the three unnatural alternations differ significantly from the probability of being attested once in a sample of 294 languages. Historical Probabilities of all three natural alternations, on the other hand, are significantly higher than the probability of being attested once in the sample (Table 5).

We can interpret these result as follows: because the Historical Probabilities of the natural alternations are significantly higher than the probability of being attested once in a sample of N languages, we expect the natural alternations to be attested more than once in a given sample of comparable size.

For the unnatural alternations, the interpretation requires further clarifications. Let us first compare the estimated 95% BC_a confidence intervals and the raw probability of being attested once in a sample. The 95% BC_a confidence intervals of all three unnatural alternations fall below the raw probability ($P(\frac{1}{294}) = 0.34\%$). The 95% BC_a CI of PND that reaches the highest probability is estimated at [0.02%, 0.11%], which is well below the raw probability of being attested once in a sample (0.34%). This observation would predict that none of the natural alternations would be attested in a given sample. The probability of being attested once in a sample, however, also bears a degree of uncertainty. Instead of comparing the 95% BC_a CIs of unnatural alternations with the raw probability, we estimate the 95% BC_a CI (based on BSC) of the difference between the probability of being attested once in a sample ($P(\frac{1}{294})$) and the Historical Probability of an alternation A_k (as proposed in (11) above). As already mentioned, none of the Historical Probabilities of the unnatural alternations in question significantly differ from the probability of being attested once in a given sample (Table 5). We can interpret this result as follows: BSC predicts that unnatural PND, IVD, and FV might or might not be attested in a given sample.

Predicting attestedness with BSC reveals another important generalization about the derivation of phonological typology within the Channel Bias approach and about typological surveys in general. When estimating the probability of being attested once in a given sample with a non-parametric bootstrap, we sample randomly with replacement from a sample that includes only one success and many more failures. This means that many bootstrap replicates (random samples with replacement) will have probability zero. For example, in one bootstrapping draw with 10,000 replicates, 3,676 replicates (or 36.8%) yielded P = 0 in estimating $P(\frac{1}{294})$. For this reason, regardless of how low the Historical Probability of an alternation A_k is, it will likely not be significantly lower than the probability of being attested once in a given sample. This generalization persists even if we estimate the difference between the two probabilities with a smoothed bootstrap. Smoothed bootstrapping is a bootstrapping technique that samples from a "kernel density estimate of the distribution" instead of sampling from the observed distribution (Wolodzko 2017). For illustrative purposes, we estimate the difference between the Historical Probability of a hypothetical alternation that requires three sound changes to arise, each of which is attested only once in a given sample, and the probability of being attested once in that sample. Let us assume our hypothetical sample size is 294. We estimate the difference using a Gaussian kernel in the *kernelboot* package (Wolodzko 2017). The difference between the two probabilities is not significant: $\Delta P_\chi = -0.33\%$ with 95% quantiles crossing zero ($[-0.01\%, 0.27\%]$).

This observation points to a problem that typological generalizations face: virtually no alternation will have Historical Probability low enough that its P_χ will be significantly lower than the probability of being attested once in a given sample. In other words, we cannot claim with great confidence for any alternation that it will not be attested in a given sample. BSC thus either predicts that some alternations will be attested in a given sample (those with P_χ higher than $P(\frac{1}{N})$) or that some alternations might or might not be attested (those for which P_χ and $P(\frac{1}{N})$ do not differ significantly). No alternation is predicted to be unattested in a given sample by the BSC. This last generalization is in fact desirable: languages feature productive synchronic alternations that require multiple sound changes, many of which are not common. One such example is Sardinian lateral sandhi (/l/ → [ʁ] / V__V; Scheer 2015) that requires at least five sound changes (l > *ɫ > w > gʷ > ɣʷ > ʁ / V__V; see Scheer 2015 and literature therein).[9]

4.4 Typological models

4.4.1 A MaxEnt model of typology

As discussed in Section 1, several proposals exist for modeling phonological learning and computationally encoding the fact that some processes are consistently underlearned. Wilson (2006) proposes a MaxEnt model (Goldwater and Johnson 2003) that differentiates prior variance σ^2 in the regularization term of different constraints. The differing degrees of learnability of some alternations result in phonological typology according to the Analytic Bias, and the differentiating σ^2 can be used to model typology from learning biases. One of the objections against this approach is that it fails to derive typological generalizations for those processes that do not show learnability differences. A growing body of work argues that while structurally complex alternations are indeed underlearned compared to structurally simpler alternations (the so-called Complexity Bias), no such biases exist when structural complexity is controlled for (the so-called Substantive Bias) (for a literature overview, see Moreton and Pater 2012a,b). In other words, many studies found no differences in learnability between natural and unnatural alternations when complexity is controlled for (Pycha et al. 2003, Wilson 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). This is especially problematic for the Analytic Bias approach to typology because there exist substantial differences

[9] The trajectory of the development of Sardinian sandhi is confirmed by related dialects that feature intermediate stages (Scheer 2015 and literature therein).

in typology between natural and unnatural alternations. For example, 15 languages of 197 surveyed (Locke 1983; reported in Hayes and Stivers 2000) feature PNV as a synchronic phonological alternation. 26 of 153 languages surveyed in Gurevich (2004; reported in Kaplan 2010) feature IVV as a synchronic alternation. To my knowledge, no systematic typological studies of final devoicing exist, but FD is one of the most frequent alternations in world's languages (Brockhaus 1995). As was already discussed in Section 4.2, the unnatural alternations PND, IVD, and FV are very rare with only PND being securely attested as a productive synchronic process (Coetzee and Pretorius 2010). The AB approach faces problems deriving this typological distribution if no differences in learnability are observed between natural and unnatural alternations.

Historical Probabilities of alternations estimated with the BSC technique offer a solution to this problem. By introducing a typological model that combines Historical Probabilities with the MaxEnt framework, we can maintain the MaxEnt approach to modeling phonological learning and differences in learnability of different alternations (Wilson 2006, Hayes and Wilson 2008) and at the same time derive typological generalizations for those processes for which no learnability biases exist.

Two different implementations of encoding learnability differences in MaxEnt models of phonological learning have been proposed: Wilson (2006) differentiates prior variance (σ^2) of constraints, whereas White (2017) differentiates prior means (μ). In both proposals, one metric becomes redundant when the other is employed. I adopt Wilson's (2006) approach of differentiating prior variance (σ^2) and claim that in a typological MaxEnt model we can encode influences of the CB on phonological typology by differentiating prior means (or, because a typological model does not involve learning, a prior Historical Weight w_χ) for different constraints.

In a MaxEnt model of phonological learning, the probability distribution over candidates is computed from the harmony of each candidate (Goldwater and Johnson 2003, Wilson 2006):

(12)

$$H(y, x) = \sum_{i=1}^{m} w_i C_i(y, x).$$

The probability of an output y given an input x ($P(y|x)$) is computed as (Goldwater and Johnson 2003, Wilson 2006):

(13)

$$P(y|x) = \frac{e^{H(x,y)}}{\sum\limits_{y \in Y(x)} e^{H(x,y)}}.$$

This paper proposes that we can extend the MaxEnt framework for modeling phonological learning to a model of *typological* probabilities. More specifically, we argue that we can model both AB and CB influences on typology in such a "typological" MaxEnt model. For modeling AB, we adopt Wilson's (2006) approach. When modeling CB influences, however, we propose that the prior Historical Weights of each Markedness-Faithfulness constraint pair for a given alternation can be directly derived from Historical Probabilities. The differences in Historical Weights between a Markedness ($\mathcal{M}$) and Faithfulness ($\mathcal{F}$) constraints for a given alternation A_k are calculated from Historical Probabilities according to:

(14)

$$\Delta w_{\chi}(\mathcal{M}, \mathcal{F}) = -\log\big(\frac{P_{\chi}(A_k)}{1 - P_{\chi}(A_k)}\big).$$

Let us look at how the MaxEnt model of typology derives typological differences between PND and PNV. The two alternations have been tested experimentally and no differences in learnability have been observed (Seidl et al. 2007). This means prior variance of the markedness *NT and *ND constraints should be kept equal. When modeling phonological learning, equal prior variance results in a desirable generalization that both alternations are equally learnable. Because of equal prior variance, we can disregard influences of the AB on typology. Without the input from Historical Probabilities, the derivation would end here and we would incorrectly predict no differences in typology between PNV and PND. Under the new model that admits both AB and CB influences, we can maintain MaxEnt framework of phonological learning and at the same time derive the observed typological differences.

To encode the CB influences on typology, we calculate the differences in Historical Weights between *NT and Ident-IO(voi) and *ND and Ident-IO constraints according to (14). Prior w_{χ} of Ident-IO constraints are set at 10.

(15) a. Δw_{χ}(*NT, Ident-IO(voi)) = 1.36
b. Δw_{χ}(*ND, Ident-IO(voi)) = 7.66

The tableau in (16) illustrates derivation of probabilities in a MaxEnt model of typology. Differences in Historical Weighs (w_{χ}) result in different harmonies ($\mathscr{H}$) of natural and unnatural candidates, which in turn results in different predicted typological probabilities (P). Predicted typological probabilities can be compared to the actual observed synchronic typology (Typol.; see Section 4.4.2). $\mathscr{H}$ is calculated according to (12), and P of each candidate according to (13). To be sure, the framework in (16) only models typology and *not* phonological learning: speakers have no access to Historical Weights.

(16)

/NT/	Ident-IO $w_{\chi} = 10$	*NT $w_{\chi} = 8.68$	$\mathscr{H}$	P	Typol.
a. [NT]		−1	−8.68	.79	.924
b. [ND]	−1		−10	.21	.076
/ND/	Ident-IO $w_{\chi} = 10$	*ND $w_{\chi} = 2.34$	$\mathscr{H}$	P	Typol.
a. [ND]		−1	−2.34	.99953	≈ .9975
b. [NT]	−1		−10	.00047	≈ .0025

The MaxEnt model of typology in (16) states that given input /NT/, the typological probability of mapping /NT/ → [ND] is 7.6%, while the typological probability of /NT/ → [NT] is 92.4%. For input /ND/, the probability of /ND/ → [NT] is 0.05%, the probability of /ND/ → [ND] is 99.9%.

The current proposal makes no claims about how differences in prior variance (σ^2) of different constraints result in observed typology. This problem is out of scope of the present paper and is left for future work. An advantage of introducing Historical Weights into a typological model within the MaxEnt framework is that such a model derives differences in the observed typology that would be left unexplained if only the AB influences would be admitted to the model.

4.4.2 Comparing P_χ to observed synchronic typology

Predictions of the BSC model can be evaluated by comparing Historical Probabilities with observed typology of synchronic alternations. Estimation of synchronic typological probabilities, however, faces many more difficulties and problematic assumptions than estimation of Historical Probabilities. 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, comparison between Historical Probabilities and observed synchronic typology can only be qualitative at this point, especially until more comprehensive and well-balanced surveys are available. Nevertheless, Historical Probabilities estimated with the BSC technique match the observed synchronic typology relatively well and, to the author's knowledge, better than alternative approaches. Table 6 compares Historical and observed synchronic probabilities. Historical Probabilities (P_χ) are estimated with the BSC technique as described above (see Section 4.1 and Table 4). The synchronic typology of natural processes is estimated based on surveys of alternations in Locke (1983) (reported in Hayes and Stivers 2000) and Gurevich (2004) (reported in Kaplan 2010). Because no systematic typologies of FD exist, this process is left out of the comparison. The synchronic typology of unnatural alternations is based on surveys in Author (2017) and Author and Name (2017). The synchronic typology of unnatural is challenging to estimate, because it is difficult to confirm a productive synchronic status for an unnatural alternation compared to natural alternations, and because typological surveys of unnatural alternations are usually not performed in a systematic and controlled manner. The surveys in Author (2017) and Author and Name (2017), for example, are based on all sources and reports available to the authors. A reasonable estimate of the languages surveyed would be approximately 600.

PND has been confirmed as a fully productive synchronic alternation in two related languages and as a morphophonological alternation in a few others. For the purpose of comparison, only fully productive alternations are counted in the synchronic typology. Because Tswana and Shekgalagari are closely related, we count PND there as a single occurrence. IVD and FV are, to the author's knowledge, not attested as productive phonological alternations in any language, which is why we estimate their synchronic typological probability below $P(\frac{1}{600})$.[10] Confidence intervals for typological probabilities of synchronic alternations are estimated with a non-parametric bootstrap (according to the same procedure described in Section 3.3.1 for estimating Historical Probabilities of alternations requiring only one sound change) from the numbers of successes (languages in a sample with a synchronic alternation) and failures (languages in a sample without the synchronic alternation).

[10] If we counted two of the best candidates for FD and IVD, Lezgian and Sula, as featuring fully productive unnatural alternations (see Yu 2004, Bloyd 2017, and Author and Name 2017), the typological probabilities of FV and IVD would be estimated at $P(\frac{1}{600}) = 0.17\%$.

Table 6: 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	14.3	10.2	18.4	7.6	4.1	11.2
PND	0.05	0.02	0.9	0.17	0.0	0.5
IVV	9.5	6.1	12.9	17.0	11.1	22.9
IVD	0.02	0.008	0.05	<0.17	0.0	0.5
FD	11.2	7.8	15.0	—		—
FV	0.01	0.004	0.03	<0.17	0.0	0.5

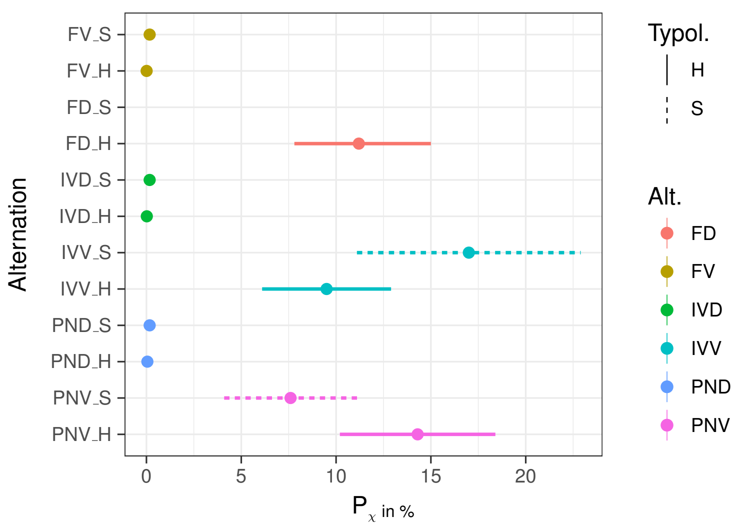

Figure 3: Observed Historical (H, solid line) and synchronic (S, dashed line) probabilities (in %) with 95% BC_a CIs from Table 6 estimated with BSC.

Table 6 and the corresponding plot of estimated Historical and synchronic probabilities with 95% BC_a CIs in Figure 3 show that BSC correctly predicts natural alternations to be significantly more frequent than the unnatural alternations. Historical Probabilities and observed synchronic typology also match to the degree that at least the 95% BC_a confidence intervals of both Historical and synchronic typological probabilities always overlap for all five processes compared. BSC also correctly predicts that some unnatural alternations, such as PND, will be more frequent than others, such as FV, which is substantiated by the observed typology. To the author's knowledge, no other approach to typology (within AB or CB frameworks) makes such predictions, especially considering the fact that we estimate Historical Probabilities of unnatural alternations not directly from observed surface typology, but from the typology of natural sound changes that are independent of the unnatural result. Given the quality of current typological surveys, this is, to the author's knowledge, the most accurate comparison between predicted Historical and observed synchronic probabilities. The comparison thus shows that Historical Probabilities estimated with the BSC technique yield an accurate prediction of synchronic typological probabilities.

5 Implications

5.1 The AB-CB conflation problem

We argue that BSC derives several typological generalizations that have so far been considered problematic for both the CB and AB approaches. Kiparsky (2006, 2008) and others (de Lacy and Kingston 2013) claim that the CB approach to typology fails to explain why some processes, such as FV or IVV, are non-existent. As was argued above, the BSC method derives these generalizations quantitatively: we predict FV and IVV to be very rare or possibly unattested in a given sample.

On the other hand, the lack of evidence in favor of the Substantive Bias (see Moreton and Pater 2012a,b) has been used as a counterargument against the AB approach to typology. If alternations show no learnability differences, substantial typological differences in these alternations are not easily derivable within the AB approach. By introducing Historical Weights derived from Historical Probabilities, we can maintain the AB approach to typology, especially for processes for which learnability differences have been confirmed experimentally, while at the same time deriving typological observations for those processes that seem to be equally learnable (Section 4.4.1). We adopt Wilson's (2006) version of encoding learnability differences with differentiating prior variance and extend the framework into a typological model that admits both AB (σ^2) and CB (w_χ) influences. We show that the BSC method directly predicts the typological prevalence of natural processes and the rarity of unnatural processes even if the two are equally learnable and no AB explanation is available.

The BSC method also offers a quantitative framework for disambiguating AB and CB influences on typology. If two typologically unequal alternations show no learnability differences, but have significantly different Historical Probabilities, we can explain differences in the observed typology between the two alternations within the CB approach. On the other hand, if two typologically unequal alternations have equal Historical Probabilities and show differences in learnability, we can explain typological differences within the AB approach. In the case of the unnatural alternations PND, IVD, and FV, the BSC technique suggests that CB primarily influences the typology (Section 4; for further disambiguation between AB and CB, see 5.2). Comparing the AB and CB influences according to this method in other alternations might yield different results.

Disambiguation and comparison of AB and CB influences on phonological typology face two challenges. First, further learnability experiments are needed to confirm the absence or presence of Substantive Bias, preferably for each alternation we are interested in. Opposing results have been reported from artificial grammar learning experiments on the existence of Substantive Bias, although rarely for the same alternation. It is entirely possible that Substantive Bias is present only in a subset of natural-unnatural alternation pairs. From an overview of the literature (Moreton and Pater 2012a,b) it appears that Substantive Bias has been more often reported for vocalic alternations than for consonantal alternations, but this observation requires further experimental evaluation. In any case, the AB influences on typology should ideally be estimated from artificial grammar learning experiments for every alternation in question.[11]

The second challenge facing the disambiguation of AB and CB influences is the fact that the outcomes and frequencies of sound change might be primarily influenced by learning biases them-

[11] Additionally, rather than deriving prior σ^2 from P-map-related metrics (in MaxEnt models of learning, Wilson 2006, White 2017), differences in learnability should be encoded directly from results of artificial grammar learning experiments. Relating differences in learnability from perceptual measures is problematic, especially because it has been recently shown in Greenwood (2016) that perceptual salience in experimental design can influence observed results. In other words, differences in learnability should ideally be estimated experimentally for every alternation from which typological generalizations are drawn.

selves (cf. Kiparsky 1995, 2008). In other words, it is possible that the AB influences crucially affect the observed typology of sound changes (let us call this the *AB-CB conflation* problem). While this is not the position taken in many other works on sound change (Labov 1994) or this paper, the accuracy of the BSC model might be undermined by the argument stating that learnability crucially influences sound change typology. Argumentation against this position is beyond the scope of this paper; as will be argued below, however, the "AB-CB conflation" problem is controlled for at least in the case of unnatural alternations.

Unnatural alternations and MSCR play a crucial role in controlling for the "AB-CB conflation" problem. Even if the probabilities of individual sound changes are crucially influenced by learnability (and therefore by AB), the fact that for unmotivated and unnatural alternations to arise, at least two or three sound changes, respectively, are required to operate in a language (due to the MSCR) means that CB plays a crucial role in determining the synchronic typological probabilities of unmotivated and unnatural alternations. All else being equal, even if we assumed learnability is the only factor influencing probabilities of individual sound changes, the probability of a single sound change will necessarily be greater than the probability of a combination of three sound changes, and this generalization is necessarily influenced by CB because the sound changes need to operate in combination and in the temporal dimension of a given speech community.

5.2 AB-CB complexity mismatch

The BSC technique proposed here identifies additional mismatches in predictions of the AB and CB approach, especially with respect to complexity of alternations and their typological attestedness. BSC not only predicts unnatural alternations will be rare (Sections 4 and 5.1), but also that, all else being equal, complex alternations will be less frequent than simple alternations. According to the "minimality principle", which states that sound change is a change in one feature in a given environment (see Section 2), 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. Exactly the same generalization is, however, also predicted by the AB 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).

Natural and unnatural alternations that arise through the Blurring Process provide unique insight into the discussion on Complexity Bias and Substantive Bias as well as the discussion of the AB-CB conflation problem. There is a crucial mismatch in predictions between the AB and CB approach with respect to unnatural alternations.

The BSC technique makes the following predictions: the more sound changes an alternation requires, the lower the Historical Probability of that alternation, regardless of its complexity (see Table 7). In other words, the BSC prediction that complex alternations will be rare is violable: if the three sound changes of a Blurring Process result in a simple unnatural alternation, the BSC still predicts 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 (MSCR). On the other hand, the AB 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 complexity of the resulting alternation, but the third sound change decreases the complexity. Complexity Bias thus predicts that the alternations that arise from the first and the second sound change 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 more frequent than the alternation requiring only two sound changes. Let us call this prediction the *AB-CB complexity mismatch*.

We can estimate Historical Probabilities for each step in the Blurring Process that leads to unnatural alternations (for a detailed description of the Blurring Process, see Section 2 and (7), (8), and (9)). Let us take as an example PND. The probability of the initial stage before the first sound change operates is calculated simply as $1-\mathrm{P}_\chi$, where P_χ is the Historical Probability of the first sound change. The Historical Probabilities of each alternation were estimated with BSC as described above.

Table 7: Mismatch in Historical Probabilities (P_χ) and probabilities predicted by complexity bias ($\mathrm{P}_{complex}$) for all four synchronic stages in a Blurring Cycle that involves three sound changes and results in a synchronic alternation PND. Historical Probabilities are estimated using BSC with 95% BC_a CI (Lo. and Up).

Sound change	Alternation	$\mathbf{P}_\chi$	Lo.	Up.	Features	$\mathbf{P}_{\mathbf{complex}}$	$\mathbf{P}_\chi$
	No alternation	67.0	61.6	72.4	0		
D > Z / [−nas]__	D → Z / [−nas]__	33.0	27.6	38.4	1	↓	↓
D > T	Z → T / [+nas]__	1.1	0.7	1.8	2	↓	↓
Z > D	PND	0.05	0.03	0.09	1	↑	↓

The fact that the first two sound changes in the Blurring Process increase the complexity of the alternations argues against the radical approach to the AB-CB conflation problem that states that sound change probabilities are primarily influenced by learnability and hence that estimated CB influences are crucially conflated with AB influences. If anything, AB 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 operating in question are not crucially influenced by AB.

The mismatched predictions of BSC and Complexity Bias illustrated in Table 7 provide crucial new information for disambiguating AB and CB biases. The AB-CB 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, the CB 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 BSC compared to structurally simpler unnatural alternations, we have a strong case in favor of the AB influence, and more precisely in favor of Complexity Bias within the AB approach to typology.

In fact, preliminary 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 BSC (and therefore CB), 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 with BSC as $\Delta\mathrm{P}_\chi(\mathrm{Z} \to \mathrm{T}\ /\ [+\mathrm{nas}]\_\_, \mathrm{PND}) = 1.1\%,\ [0.6\%, 1.7\%]$. The Historical Probability of the alternation Z → T / [+nas]__ that arises through two sound changes is thus predicted to be approximately 20 times more frequent than the Historical Probability of PND (see Table 7). 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, Nasioi, and Punu (Hyman 2001, Merrill 2014, 2016a,b, Santos 1996, Brown 2017).[12] Other languages are more difficult to classify because some of them seem to have 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 to be smaller than predicted by BSC.

What is even more intriguing is the high frequency at which the third sound change in the Blurring Process, occlusion of voiced fricatives to stops, operates on synchronic systems that have 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 is $P_\chi = 12.5\%$, $[7.9\%, 17.1\%]$. Of the languages in the survey in Author (2017) that undergo the first two sound changes that lead to PND, 6 languages (out of 7, or approximately 85.7%)[13] feature occlusion of stops for at least one place of articulation or in at least in one position in the word. If we count only cases in which occlusion of fricatives targets more than two places of articulation, only Tswana, Shekgalagari, Makuwa, and Murik would count (two independent changes). 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 BSC (and therefore CB) predicts for the occlusion of voiced fricatives in general.

We can compare the Historical Probability of the occlusion of fricatives based on the surveys of sound changes in Kümmel (2007), and the observed probability of occlusion of fricatives in those languages that have already undergone the first two sound changes in the Blurring Cycle that leads to PND. Because both of these estimations involve a single sound change and because the second sample is small (7 observations), we test the significance of the difference with Fisher's Exact Test. The number of languages with voiced fricatives (216 surveyed) that undergo occlusion of voiced fricatives is 27. As already mentioned, under the less conservative count, 6 out of 7 languages in the Blurring Cycle show occlusion for at least one place of articulation or at least for one context (word-initially in Nasioi). The difference between the two counts is statistically significant ($p < 0.0001$).

This suggests that the high occurrence of the third sound change in the Blurring Process (in the case of PND, the occlusion of fricatives), might be an influence of the Complexity Bias within the AB approach. While AB likely does not influence the probabilities of the first two sound changes in the Blurring Process, because they increase complexity and therefore lower the learnability, it is likely that the occurrence of the third sound change and the therefore lower probability of the more complex unmotivated alternation is influenced precisely by Complexity Bias.

This example shows that mismatches in predictions between AB and CB identified by the Blurring Process, MSCR, and BSC can shed new light on the discussion of AB vs. CB influences on typology. Instead of arguing in favor of or against one or the other approach, influences of AB and CB should be estimated quantitatively. Comparison of quantitative estimates can provide new information on what aspects of typology are primarily influenced by AB and what aspects primarily by CB. This paper proposes the BSC technique, a quantitative method for estimating CB influences. For every alternation, BSC offers a technique for estimating its Historical Probability, i.e. a CB contribution to the observed typology. Estimation of AB influences should involve further experimental data: ideally, we can test the learnability of every alternation and compare it to its Historical Probability estimated with BSC. Further and more precise estimations of AB influences

[12] Punu is a language that undergoes a different development for PND from the one described in this paper. For a discussion, see Hyman (2001).

[13] PND in Tswana, Shekgalagari, and Makhuwa are counted as one occurrence. South Italian dialects that devoice affricates are not counted. We also exclude Mpongwe from the count because of the limited description and marginal status of PND there.

with experimental data and application of the BSC method to novel unnatural alternations should yield further insights into the long-standing discussion in phonology.

6 Conclusion

This paper presents applications of a technique for estimating Historical Probabilities of alternations called Bootstrapping Sound Changes (first outlined in Author 2017). We give a detailed description of the statistical model and discuss its assumptions, properties of the sample, and implementation. The paper also includes functions in the R Statistical Software language for performing the BSC analysis.

Several applications of the BSC technique are presented. For any synchronic alternation, both attested and unattested, the BSC technique estimates its Historical Probability from the number of sound changes the alternation requires and their respective probabilities. In other words, the BSC technique quantifies predictions of the Channel Bias approach to typology that can be compared to the actual observed synchronic typology. BSC also allows for inferential statistical tests comparing the Historical Probabilities of any two alternations. The BSC technique additionally predicts (un)attestedness of an alternation in a given sample. Alternations are either predicted to be attested more than once in a given sample or we predict an alternation to be possibly attested or unattested. The paper also discusses how Historical Probabilities estimated with BSC can be used in a MaxEnt model of typology, which adopts the MaxEnt framework of phonological learning and at the same time predicts typological rarity of alternations that do not show differences in learnability based on experimental data.

Finally, we argue that comparing Historical Probabilities with the observed typology yields new insights into the discussion about AB vs. CB influences on typology. First, we show that BSC predicts the observed typology with relatively high accuracy. This is especially true for differences between natural and unnatural alternations, which pose a problem for the AB approach to typology. To the author's knowledge, BSC makes the most accurate typological predictions currently possible within the CB approach. For example, no other proposals known to the author predict with significance that unnatural alternations will be substantially less frequent than natural alternations and at the same time predicts some unnatural alternations will be significantly less frequent than others, a situation that is substantiated by the observed typology. Finally, BSC identifies crucial mismatches in predictions between AB and CB that provide new information for disambiguating AB vs. CB influences on typology. Both AB and CB predict that complex alternations will be less frequent, but within CB this prediction can be violated in the case of unnatural vs. unmotivated alternations. Based on these mismatches, we argue that the typological difference between natural and unnatural alternations is primarily due to the Channel Bias, but that the relatively high frequency of unnatural alternations compared to complex unmotivated processes is due primarily to Complexity Bias.

These predictions have further theoretical implications. Synchronic grammar should ideally derive all observed patterns while at the same time excluding impossible processes. Typological observations often prompt adjustment in grammar design. Our proposed framework suggests that some typological gaps are historical accidents which need not be encoded in synchronic grammars and quantifies these gaps. Estimation of the CB and AB 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 observations from diachronic development. Our proposed model hopes to provide a step in this direction.

In sum, this paper aims to shift the discussion in phonological typology from arguing in favor or against the Analytic Bias and Channel Bias approach to estimating both influences quantitatively. Application of the BSC technique on new alternations combined with new experimental data has the potential to yield further information for this long-standing discussion in phonology.

References

Adelaar, Willem F. H. 1977. *Tarma Quechua: Grammar, Texts, Dictionary.* Lisse: The Peter de Ridder Press.

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.

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

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

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

Brockhaus, Wiebke. 1995. *Final Obstruent Devoicing in the Phonology of German.* Tübingen: Niemeyer.

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.

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.

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.

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 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.

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.

Goldwater, Sharon and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson and Östen Dahl (eds.), *Proceedings of the Workshop on Variation within Optimality Theory.* 111-120. Stockholm University.

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.

Gurevich, Naomi. 2004. *Lenition and Contrast: The Functional Consequences of Certain Phonetically Conditioned Sound Changes*. New York: Routledge.

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

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.

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

Hayes, Bruce, and Tanya Stivers. 2000. *Postnasal Voicing*. Ms. Available Jun 6, 2015 at: `www.sscnet.ucla.edu/soc/faculty/stivers/TS_website/Publications_files/Postnasal_voicing.pdf`

Hayes, Bruce and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3): 379-440.

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.

Janson, Tore. 1991/1992. Southern Bantu and Makua. *Sprache und Geschichte in Afrika* 12/13: 1-44.

Janssens, Baudouin. 1993. Doubles réflexes consonantiques: quatre études sur le bantou de zone A (bubi, nen, bafia, ewondo). Ph.D. Dissertation, Université libre de Bruxelles, Faculté de Philosophie et Lettres, Bruxelles.

Kaplan, Abby. 2010. Phonology shaped by phonetics: The case of intervocalic lenition. Ph.D. Dissertation, University of California, Santa Cruz.

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.

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.

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

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.

Locke, John. 1983. *Phonological Acquisition and Change*. New York: Academic Press.

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.

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

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.

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.

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.

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

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.

Rohlfs, Gerhard. 1949. Lautlehre. Vol. 1 of *Historische Grammatik der Italienischen Sprache und ihrer Mundarten*. Bern: Francke.

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

Scheer, Tobias. How diachronic is synchronic grammar? Crazy rules, regularity, and naturalness. In *The Oxford Handbook of Historical Phonology*, edited by Patrick Honeybone and Joseph Salmons. Oxford: Oxford University Press,

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.

Solé, Maria-Josep, Larry M. Hyman, and Kemmonye C. Monaka. 2010. More on post-nasal devoicing: The case of Shekgalagari. *Journal of Phonetics* 38(4): 299-319.

Stanton, Juliet. Learnability shapes typology: The case of the midpoint pathology. *Language* 92(4): 753–791.

Staubs, Robert. 2014. Computational modeling of learning biases in stress typology. Ph.D. Dissertation, University of Massachusetts Amherst.

Westbury, John R. and Patricia A. Keating 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22(1): 145-166.

White, James. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language* 93(1): 1-36.

Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer.

Wilson, Colin. 2003. Experimental investigation of phonological naturalness. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, edited by Gina Garding and Mimu Tsujimura. 533-546. Somerville, MA: Cascadilla Press.

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

Wolodzko, Tymoteusz. 2017. kernelboot: Smoothed Bootstrap and Random Generation from Kernel Densities. R package version 0.1.2. url: CRAN.R-project.org/package=kernelboot

Xromov, Al'bert. 1972. *Jagnobskij jazyk.* Moscow: Nauka.

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

Nov 2017

A Supplementary materials

A.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 A.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)))
  )
  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)
}

# Example:
pnd.counts <- c(97, 18, 27)
pnd.surveyed <- c(294, 263, 216)

pnd <- bsc(pnd.counts, pnd.surveyed)
summary.bsc(pnd)

# Output:
```

```
##BOOTSTRAPPING SOUND CHANGES
##
##Observed P = 0.04704 %
##Estimated 95 % BCa CI = [ 0.0261 %, 0.0862 %]
```

A.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 A.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)
}

# Example:
pnd.counts <- c(97,18,27)
pnd.surveyed <- c(294,263,216)
```

```

fv.counts <- c(6,32,27)
fv.surveyed <- c(294,294,88)

pnd <- bsc(pnd.counts, pnd.surveyed)
fv <- bsc(fv.counts, fv.surveyed)

pndfv <- bsc2(pnd, fv)
summary.bsc2(pndfv)

#Output:
##BOOTSTRAPPING SOUND CHANGES - COMPARE
##
##Observed Delta P = 0.03568 %
##Estimated 95 % BCa CI = [ 0.0114 %, 0.0744 %]
##
##P(A1) is significantly higher than P(A2).
```

A.3 summary.bsc()

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

A.4 summary.bsc2()

The function *summary.bsc2()* computes the 95% BC_a CI for the bootstrap replicates based on the *bsc2()* function (see A.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 A.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")
}
```

A.5 plot.bsc()

The function *plot.bsc()* takes the output of *bsc()* as input and plots the distribution of bootstrap replicates with the observed Historical Probability of the process (solid line) and 95% BC_a CI (dashed line), calculated with the *boot.ci()* function from the *boot* package. The plotting is based on the *ggplot2* package (Wickham 2009). An optional argument *Alternation* allows for the change of the name of the alternation in the legend.

```
plot.bsc <- function (bsc.alt, Alternation = c("Alternation")) {
  library(ggplot2)
  bsc.ci.alt <- boot.ci(bsc.alt, type = "bca")
  bsc.alt.df <- data.frame(bsc.alt$t)
  bsc.alt.df$name <- Alternation
  names(bsc.alt.df) <- c("boot", "Alternation")
  boot.plot <- ggplot(bsc.alt.df, aes(boot, fill = Alternation)) + geom_density(
      alpha = 0.5)  +
    geom_vline(xintercept = bsc.alt$t0, colour="red",linetype = "solid") +
    geom_vline(xintercept = bsc.ci.alt$bca[4],
               colour="red", linetype = "dashed") +
    geom_vline(xintercept = bsc.ci.alt$bca[5],
               colour="red", linetype = "dashed") +
    theme_bw() + xlab("Px␣in␣%") + ylab("")
  return(boot.plot)
}

# Example:
pnd.counts <- c(97, 18, 27)
pnd.surveyed <- c(294, 263, 216)

pnd <- bsc(pnd.counts, pnd.surveyed)
plot.bsc(pnd, alternation = "PND")
```

A.6 plot.bsc2()

The function *plot.bsc2()* takes the output of *bsc()* as its input (two alternations) and plots the distribution of bootstrap replicates with the observed Historical Probability of the process (solid line) and 95% BC_a CI (dashed line), calculated with the *boot.ci()* function from the *boot* package for each alternation. The plotting is based on the *ggplot2* package (Wickham 2009). An optional argument *Alternation* allows for the change of the name of the two alternations in the legend. Note that *plot.bsc2()* does not plot bootstrap replicates of the difference between two Historical Probabilities, but rather bootstrap replicates of Historical Probabilities of each of the two alternations. To plot the bootstrap replicates of the difference between two Historical Probabilities, apply *plot.bsc()* to the output of *bsc2*.

```
plot.bsc2 <- function (bsc.alt1, bsc.alt2, Alternation = c("Alternation␣1","
    Alternation␣2")) {
  library(ggplot2)
  bsc.ci.alt1 <- boot.ci(bsc.alt1, type = "bca")
```

```
  bsc.ci.alt2 <- boot.ci(bsc.alt2, type = "bca")
  bsc.alt1.df <- data.frame(bsc.alt1$t)
  bsc.alt2.df <- data.frame(bsc.alt2$t)
  bsc.alt1.df$name <- Alternation[1]
  bsc.alt2.df$name <- Alternation[2]
  names(bsc.alt1.df) <- c("boot", "Alternation")
  names(bsc.alt2.df) <- c("boot", "Alternation")
  bsc.alt.df<-rbind(bsc.alt1.df, bsc.alt2.df)
  boot.plot <- ggplot(bsc.alt.df, aes(boot, fill = Alternation)) +
    geom_density(alpha = 0.5)  +
    geom_vline(xintercept = bsc.ci.alt1$bca[4],
               colour = "red", linetype = "dashed") +
    geom_vline(xintercept = bsc.ci.alt1$bca[5],
               colour = "red", linetype = "dashed") +
    geom_vline(xintercept = bsc.alt1$t0,
               colour = "red", linetype = "solid") +
    geom_vline(xintercept = bsc.ci.alt2$bca[4],
               colour = "turquoise4", linetype = "dashed") +
    geom_vline(xintercept = bsc.ci.alt2$bca[5],
               colour = "turquoise4", linetype = "dashed") +
    geom_vline(xintercept = bsc.alt2$t0,
               colour = "turquoise4", linetype = "solid") +
    theme_bw() + xlab("Px in %") + ylab("")
  return(boot.plot)
}

#Example:
pnd.counts <- c(97,18,27)
pnd.surveyed <-c (294,263,216)

fv.counts <- c(6,32,27)
fv.surveyed <- c(294,294,88)

pnd <- bsc(pnd.counts, pnd.surveyed)
fv <- bsc(fv.counts, fv.surveyed)

plot.bsc2(pndfv)
```