Post-Nasal Devoicing and a Probabilistic Model of Phonological Typology

Gašper Beguš
begus@fas.harvard.edu

Abstract

This paper addresses one of the most contested issues in phonology: the derivation of phonological typology. I argue that the unnatural process of post-nasal devoicing does not derive from a single unnatural sound change, but rather from a combination of two or three sounds changes, each of which is natural and well-motivated. Evidence from both historical and dialectal data is brought to bear to create a model for explaining present and future instances of apparent unnatural alternations. Based on these findings, I propose a new probabilistic model for explaining phonological typology within the channel bias approach: I provide a proof establishing the minimal sound changes required for an unmotivated/unnatural process to arise. Finally, the paper sets grounds for calculating “historical probabilities” of phonological alternations. This approach is shown to encounter no major difficulties in explaining phonological typology, contrary to what has been claimed in the literature.

Keywords: phonological typology, sound change, naturalness in phonology, probabilistic model

1 Introduction

One of the most contested issues in phonology concerns whether typological patterns found in phonological data result from constraints in the innate grammar and speakers’ cognitive predispositions (UG) or from channel of transmission (constraints on sound change). Two opposing proposals emerge from this discussion: the analytic bias hypothesis and the channel bias hypothesis (Moreton 2008). The analytic bias approach assumes that typological patterns emerge because innate cognitive biases rule out certain phonological processes (cf. Kiparsky 2006, 2008). Channel bias, on the other hand, assumes that constraints on sound change are responsible for typological patterns: an inherent directionality of sound changes results in a predictable pattern of phonetic processes and phonologizations that ultimately determine surface typology (cf. Hyman and Schuh 1974, Ohala 1993, Bybee 2001, Blevins 2004, 2006, Yu 2012, see also Hansson 2008 for a survey).¹

¹ These two approaches have also been given other labels, such as “amphichronic” vs. “evolutionary” phonology (Blevins 2004, 2006, Kiparsky 2008).
An adequate phonological theory must, at minimum, predict attested grammars and rule out unattested (or impossible) ones. Any conclusive evidence in favor of either the analytical or the channel bias approach is difficult to reach primarily because extant typological patterns can be derived more or less equally well by both approaches, which also face similar challenges. Take an example from Moreton (2008), who shows that vowel-height harmony is common, whereas consonant-continuacy harmony is non-existent. Analytic bias accounts for this pattern by positing that UG contains the constraint \textit{AGREE}(high), but not \textit{AGREE}(continuant). The learner is thus unable to acquire consonant-continuacy harmony or will acquire it at a much slower (lexicalized) pace, precisely because of the constraints on UG. The channel bias, on the other hand, explains this typological discrepancy by assuming the phonetic viability of a process of subtle variation in vowel height of two consecutive vowels due to coarticulation. Most of the time, speakers are able to recover the “intended phonological height” of a phoneme, but when this recovery fails, the phonetic process becomes a phonological process in the hearer’s grammar. Consonant-continuacy harmony is ruled out by the fact that there is no “phonetic precursor” for variation and the phonetic process that would lead to phonologization is therefore not possible (Moreton 2008:86).

That both hypotheses can derive surface typology equally well is actually not surprising, since both approaches rely heavily on surface typology to begin with. The analytical bias approach begins with typological generalizations and sets up its constraint inventory accordingly. The channel bias approach seeks to explain the phonetics behind phonological processes and allows or disallows certain phonetic processes based on typology too. It is not the goal of this paper to discuss the explanatory power of one approach over the other. Rather, the aim of this study is to describe and explicate an exemplar phenomenon that bears implications for the contest between the two approaches: post-nasal devoicing (PND).

Because the two approaches are arguably equally powerful in explaining typological generalizations, the typology itself will not yield decisive arguments in favor of either proposal; thus, we must look for decisive evidence elsewhere. The discussion has recently turned to experimental evidence (for the discussion, see Zuraw 2007, Moreton and Pater 2012, Yu 2012, Hayes and White 2013, White 2013), but here, too, no firm conclusions have yet been drawn. In this paper, I will focus on one salient vector along which the two approaches crucially differ: the degree to which they constrain either innate grammar or sound change. As Kiparsky (2006) points out, the analytic bias approach allows sound change to be “fairly unconstrained”: the learner’s grammar is the component that filters impossible alternations. Sound change can thus, in principle, operate in any direction. As Kiparsky (2008) notes: “Whatever change can create, it can also destroy” — in other words, it’s the structure that constrains change, and not vice-versa (Kiparsky 2008). Channel bias, on the other hand, arrives at typological predictions by restricting the operation of sound change: sound change can only operate in certain (phonetically justified) directions and is assumed to be impossible in other (non-justified) directions. This restriction on directionality of sound change (Greenberg 1966, Hyman and Schuh 1974, Bybee 2001, Blevins 2004, Hyman 2008) results in phonological typology.

---

2 Ideally, phonetic viability is supported by empirical data; however, often it is only assumed based on typological data.
The two approaches are not mutually exclusive and it is likely that both factors play a role in shaping phonological typology. This position is being accepted by an increasing number of scholars (cf. Moreton 2008, Hayes and White 2013). However, our understanding of how exactly the two factors interplay and shape phonology is still incomplete, primarily because elaborate models of the emergence of phonological typology are lacking on both sides of the spectrum. One of the goals of this paper is to disambiguate the two approaches, provide argumentation for the existence of channel bias, and present a detailed account of how exactly channel bias shapes phonology – a question that has been regarded as crucially problematic within the channel bias approach. This should bear a more general implication for the discussion on how the two major factors interact to shape phonological typology.

1.1 Background

As mentioned, both approaches face challenges, one of the biggest of which is how to account for typologically unusual (and typically rare) alternations. Probably the most thoroughly discussed typological oddity on both sides of the analytical spectrum is word-final or coda voicing (T → D / _#). Word-final voicing is assumed to be highly unnatural process that is either impossible or unattested as a synchronic phonological process. If this process were to exist and persist synchronically, it would operate against a universal phonetic tendency of coda devoicing (D → T / _#). Voicing is generally a difficult process to maintain, and this difficulty is only enhanced word-finally, where stops are produced “with reduced pulmonary pressure” (Iverson and Simon 2011: 1633, Blevins 2004) and the voicing process loses its cues (Steriade 1997, Iverson and Simon 2011). Indeed, the pressure in the opposite direction from such a process is so great that phonetic devoicing is expected even in languages without phonological devoicing. Word-final devoicing fits the bill for a universal phonetic tendency: (a) it has a well-motivated phonetic explanation; (b) there exists a phonetic tendency to devoice final stops even in languages without phonological devoicing; (c) it is very common and well-attested cross-linguistically. Thus, any putative process of word-final voicing would necessarily operate against a universal phonetic tendency.

Proponents of the analytic bias hypothesis assume word-final voicing to be non-existent in synchronic grammar cross-linguistically. For instance, Kiparsky (2006) argues that the apparent cases of final voicing reported in Blevins (2004) do not to qualify as cases of synchronic final voicing — or, at least, they have competing alternative explanations. A more robust example of coda voicing is found in Lezgian (Yu 2004, Haspelmath 1993). Yu (2004:77) reports that Lezgian distinguishes four stop series prevocally (plain voiced, voiceless ejective, voiceless aspirated, and plain voiceless), which in coda position get reduced to a three-way distinction: the plain voiceless series and voiced series merge into a single voiced series. The distinction is illustrated in (1).

(1) Lezgian (from Yu 2004: 77)

\[
\begin{array}{ccc}
\text{V} & \bar{I}_o \\
/D/ & D & D \\
/T^*/ & T^* & T^* \\
/T^h/ & T^h & T^h \\
/T/ & T & D \\
\end{array}
\]
The merger of the plain voiceless and voiced series into a voiced series in coda position indicates that Lezgian has a synchronic phonological process of coda voicing ($T \rightarrow D / \sigma$), which is limited to monosyllabic words. The examples in (2) (from Yu 2004) illustrate this process.

(2)

<table>
<thead>
<tr>
<th>V</th>
<th>$\sigma$</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>pab</td>
<td>pap-a</td>
<td>‘wife’</td>
</tr>
<tr>
<td>rab</td>
<td>rap-uni</td>
<td>‘needle’</td>
</tr>
<tr>
<td>seb</td>
<td>sep-erar</td>
<td>‘curse’</td>
</tr>
<tr>
<td>rad</td>
<td>rat-uni</td>
<td>‘intestine’</td>
</tr>
<tr>
<td>gad</td>
<td>gat-u</td>
<td>‘summer’</td>
</tr>
<tr>
<td>pad</td>
<td>pat-ar</td>
<td>‘side’</td>
</tr>
<tr>
<td>leg$^w$</td>
<td>lek$^w$e</td>
<td>‘tub’</td>
</tr>
<tr>
<td>rug$^w$</td>
<td>ruk$^w$adi</td>
<td>‘dust’</td>
</tr>
<tr>
<td>pag$^w$</td>
<td>pak$^w$ar</td>
<td>‘rib’</td>
</tr>
</tbody>
</table>

The phonetic study in Yu (2004), however, shows that underlying voiced and plain voiceless series do not neutralize completely: there exists a statistically significant difference between the outcome of the two series in word-final position. Voiced consonants that derive from underlying plain voiceless stops have a significantly longer closure duration as well as a longer duration of voicing into closure. If we wish to maintain that Lezgian voices coda consonants, we must, at the same time, assume that these consonants receive (at least phonetic) lengthening as well. It is unclear from a synchronic perspective why this should happen.

The fact that the two series do not neutralize completely allows Kiparsky (2006) to propose an alternative analysis. He assumes that the Lezgian synchronic phonological system has four series of stops — but unlike Yu, he proposes that the fourth series consists of voiced geminates. Thus, instead of coda voicing, he assumes that the process in (2) consists of onset degemination and devoicing. His analysis is illustrated in (3).

(3) Lezgian (from Kiparsky 2006)

$$\begin{array}{c|c|c|c}
\text{V} & \text{ } & \text{$\sigma$} \\
/D/ & D & D \\
/T'/$ & T' & T' \\
/T^h/ & T^h & T^h \\
/D:/ & T & D:
\end{array}$$

This analysis, too, has its shortcomings: like Yu, Kiparsky has to devise a two-step process — devoicing and degemination of voiced geminates in onset position — and onset devoicing is not a particularly common process in its own right. However, this derivation is by no means impossible, and Kiparsky (2008) provides evidence from other languages such as Mordva, Ewondo, and Lac Simon Algonquian (Iverson 1983) demonstrating that initial devoicing is a...
possible synchronic phonological process. As a sound change, such development may be attested in Anatolian and in Selkup of the Samoyedic group (Kümmel 2007).

Interestingly, the historical background for Lezgian alternation provided by Yu (2004) better reflects Kiparsky’s degemination analysis than his own final voicing analysis. The alternation in Lezgian arises, according to Yu (2004), through a combination of sound changes (pretonic gemination; stop devoicing; degemination) and stress assignment. Stress in Lezgian surfaces on the second syllable of non-monosyllabic words, while stress on monosyllabic words occurs on the only syllable; in the latter case, word-final voiced stops remain unchanged. In suffixed forms, however, the stress shifts, causing the underlying voiced stop to become pretonic and therefore devoiced, geminated, and finally degeminated. The development is illustrated in (4) (from Yu 2004: 87).

\[
\begin{align*}
\text{CVD} + V & > CVTT \text{ } \text{ } V > CVT \text{ } \text{ } V \\
\text{e.g.} & \quad *'d'yd \quad > \quad *'tyd \quad > \quad tyd \quad ‘throat’ \\
& \quad *d'y'd-y \quad > \quad *ty't't-y \quad > \quad tyty
\end{align*}
\]

Although Lezgian provides an apparently compelling example of final voicing, this analysis encounters two major problems. First, the voicing process does not apply in all environments, but only in a subset of monosyllabic words. Second, the plain voiceless and voiced series do not neutralize completely in coda position; a phonetic difference remains detectable between the two series. These problems pave the way for alternative proposals that analyze the alternating series as underlyingly voiced and assume that the synchronic phonological process in Lezgian is in fact onset devoicing rather than final voicing. Additionally, a lack of speaker data from nonce-word tests makes it difficult to determine how productive this process actually is.

Kiparsky (2006, 2008) concludes that coda or final voicing is an impossible phonological process and is therefore unattested, because it would operate against UG. If this is so, then sound change appears to be less constrained than synchronic grammar in at least one respect — final devoicing — since Kiparsky correctly points out that there are several potential historical scenarios through which final voicing could have arisen (he lists five in total). For example, one potential scenario would produce final voicing from a combination of intervocalic lenition and degemination. If voiceless stops were to undergo postvocalic voicing followed by degemination, the resulting synchronic system would exhibit final voicing.

One possible scenario for final voicing (from Kiparsky 2006)

\[
\begin{array}{ccc}
\text{*T} & \text{*T} & \text{*TT} \\
\text{stage 1} & \text{tat} & \text{tata} \\
\text{lenition} & \text{tad} & \text{tada} \\
\text{degemination} & \text{tad} & \text{tada} \\
\end{array}
\]
The fact that final voicing is unattested is often offered as proof that the speaker’s grammar filters impossible systems. To maintain this position, we must acknowledge that sound change is less constrained than learners’ grammar: it is capable of producing final devoicing, but internal biases in speaker’s mental grammar rule out such alternations. In fact, Kiparsky (2008) goes even further and suggest that “changes that subvert universals must either be blocked, or the system they appear to give rise to must be reanalyzed.” In other words, any potential sound change that would cause the system to operate against UG would be blocked.

The non-existence of final voicing speaks in favor of the analytic bias approach, since it suggests that UG is more constrained than sound change. In further favor of the analytic bias approach, Kiparsky (2006) claims that the channel approach has difficulty explaining why systems like the one seen in Lezgian are impossible. In this paper, I show that no such problem exists, and I propose a model for deriving surface typology within the channel bias approach.

1.2 Aims

Before we turn to the focal case of this paper, through which I aim to provide insight into the channel bias/analytic bias debate, it is first necessary to clarify the distinction between a single sound change and a combination of sound changes. A (single instance of a) sound change constitutes a change from A to B in an environment X, as a result of which A differs from B in one feature; a combination of sound changes is a set of such individual sound changes. I posit that there is no limit to how many combinations of sound changes can occur during a particular period of a language’s history, nor even to the order in which they occur. This stipulation might seem very obvious, but it is important to make this distinction explicit, especially in preparation for our discussion on the relative constraints on synchronic grammar vs. sound change. The use of the term “sound change” in the literature is often confusing and fluctuates between denoting a single instance of sound change and a combination of sound changes. I will strictly distinguish between these two notions in the remainder of this paper.

The issue of post-nasal devoicing (PND) sheds useful light on the analytic/channel bias discussion for several reasons. Just like final voicing, PND operates against a universal phonetic tendency: in this case, the tendency to voice post-nasal stops. Just like final voicing, the inverse of PND — post-nasal voicing (PNV) — is phonetically well motivated and very common cross-linguistically. However, unlike final devoicing, PND is attested as a synchronic phonological process in Tswana and Shekgalagari. Neutralization is shown to be complete in the case of PND (Coetze and Pretorius 2010), and no alternative analyses are forthcoming. Moreover, the nonce-word test presented in Coetze and Pretorius (2010) shows that PND is a productive process and not lexicalized or unproductive in any other respect.

It is, however, not entirely clear how PND came into being. For Tswana and Shekgalagari, where PND is attested as a synchronic process, various hypotheses have been offered as to how the system arose. Outside of this context, PND as a sound change has been proposed in six additional languages from three different language families. For these cases, too, several different explanations have been offered, ranging from proposals that assume PND operated as a single instance of sound change (Blust 2005, Solé et al. 2010; Solé 2012) to proposals that analyze PND as the result of a combination of sound changes (Dickens 1984, Hyman 2001).
If PND were indeed found to operate as a single sound change (as is proposed in the literature), it would constitute a rare example of sound change operating against a strong and universal phonetic tendency. This would, in turn, provide evidence in favor of the analytic bias approach, which has the freedom to assume sound change to be fairly unconstrained and places the burden of constraining the typology on the learner’s grammar. If PND were found not to operate as a sound change, we would have the opposite situation: it would seem that sound change is more constrained than synchronic grammar. If we can show that one of the rare reported unnatural sound changes is in fact a product of a combination of natural sound changes, this will lend support to the position that sound change cannot operate against universal phonetic tendency. Such a finding would indicate that, although sound change is restricted by naturalness, unnatural alternations can nevertheless arise in synchronic phonological grammars. This finding will be crucial for establishing a model that will derive typology within the channel bias approach.

The existence of so many opposing explanations of PND’s origins stems from the fact that all cases of PND have so far been studied in isolation. The aims of this paper are twofold: to address the existence of PND devoicing as a sound change, and to discuss the implications of these results for our broader understanding of phonological typology. First, I discuss all known cases of PND, identifying common patterns among those cases that provide strong evidence against analyses that invoke PND as a unitary sound change. I show that PND did not operate as a single sound change in any of the eight languages in which it is attested. Instead, I show that, in each case, a set of two or three sound changes occurring in tandem gave rise to apparent, but not actual, PND; I propose a model for explaining such “unnatural” alternations in the future. This study also has implications for the discussion of the need for sound change to be natural and phonetically motivated (for an overview of this discussion, see Blust 2005, Blevins 2008a, 2008b, Wedel 2009). On the basis of PND — a rare reported case of unnatural sound change that turns out to be a result of combination of natural sound changes — I argue that a single instance of sound change does strictly follow phonetic tendencies, is always phonetically natural and motivated, and is more constrained than synchronic grammar.3

This final point is clear from the fact that PND is not attested as a single sound change, but, as is shown in Coetzee and Pretorius (2010), is attested as a single synchronic process. The existence of at least one clear example where a single instance of sound change is more constrained than synchronic grammar provides evidence against the radical view of the analytic bias approach, which constrains the synchronic grammar and leaves sound change unconstrained. In the final section of this paper, I will capitalize on the fact that sound change is more constrained than synchronic grammar to develop a model that accounts for phonological typology based on a purely external factor: the relative probability of sound change occurrences.

On the basis of the seven examples of PND described in the literature (and an additional one that has been discovered recently), I argue that combinations of sound changes, as opposed to single instances of a sound change, are unconstrained. As it appears from the described cases,

---

3 One could argue that a single instance of sound change can produce PND, but that the alternation would be unlearnable and would never surface in language. Contra this possibility, I will show that PND is learnable and was in fact incorporated into the synchronic system of Tswana and Shekgalagari as a productive synchronic alternation. Such an argument is therefore impossible to maintain.
any number of natural sound changes can operate one after another in a given time frame, with no restrictions on the nature or order of these sound changes. This means that the final sound change in a series of sound changes does not get blocked by UG (pace Kiparsky 2008) even though it causes unnatural alternation to arise. I further show that unnatural alternations operating against universal phonetic tendencies can become incorporated productively into synchronic phonological grammars via such sound changes.

Similar positions have already been held in the literature: “telescoping,” for example, describes a phenomenon in which a sound change A → B in the environment X is followed by B → C, resulting in a sound change A → C that may not be phonetically motivated in environment X (Wang 1968, Kenstowicz and Kisseberth 1977: 64, cf. also Stausland Johnsen 2012). This paper takes the concept of telescoping one step further, by focusing on alternations that are not only unmotivated, but that operate in exactly the opposite direction from universal phonetic tendencies. I will make a distinction between “unmotivated” and “unnatural” sound changes throughout the paper (for similar distinctions, see Blevins 2004, 2008, Morley 2012): the former lack phonetic motivation, while the latter operate against universal phonetic tendency.

The main contribution of the paper is thus the model I introduce for deriving phonological typology. If it is true that combinations of sound changes are unconstrained, we still have to explain why some alternations are rare or non-existent. In fact, Kiparsky (2006) claims that one of the major shortcomings of the channel bias approach is that it cannot derive surface typology and rule out “impossible” alternations. This paper shows that this is false: in the last section, I develop a model for explaining why phenomena like PND and final voicing are typologically rare or even unattested without relying heavily on the constraints of synchronic grammar. The new model of typology proposed in this paper not only derives the fact that some processes are rare, but also provides a clear explanation for why some processes are rarer than others (or even unattested). Unlike previous proposals, the model establishes the minimal number of sound changes needed for an unnatural or unmotivated process to arise; additionally, by restricting the operation of this set of sound changes to a time frame t, the new model enables better predictions than existing models. I provide a proof to show that, for an unnatural process to arise, at least three sound changes are needed. This new model shows that even fine-grained surface typology can easily be derived within the channel bias approach.

The structure of the paper is as follows: in section 2, I discuss PND as a sound change. I bring together all eight reported cases, point to common patterns, and argue that in each of these cases, PND did not operate as a single sound change, but arose as a result of a combination of

---

4 Some scholars might object that combinations of sound changes need to be restricted as well, in order to avoid the emergence of phonological processes or systems that are impossible in human language. Restriction against unrepresentable systems, however, follows automatically from the sound change restriction proposed in this paper: articulatory and perceptual limitations are already encoded in the restriction against unnatural single sound change itself. Any single instance of sound change is defined as natural, and therefore phonetic precursors are required. The probability and naturalness of a given sound change crucially relies on the phonemic inventory of the language in question. Furthermore, all sound changes have to be active within a given time frame t in order to function in combination. We will see that even a combination of three sound changes is rare: the additional factor that restricts sound change combinations is thus a temporal one.

5 A more conservative count gives six cases of PND; some scholars count Sicilian and Calabrian and Tswana and Shekgalagari as two cases rather than four, due to the close genetic relationships between these language pairs.
two or three separate sound changes, each of which was independently phonetically motivated. This section also includes a discussion of previous proposals. In section 3, I discuss PND as a synchronic phonological process and summarize the main findings from Coetzee et al. (2007) and Coetzee and Pretorius (2010), which show that PND in Tswana is a productive phonological process. In section 4, I propose a new model for explaining typological patterns in phonology on the basis of the historical and synchronic analysis of PND.

2 Post-nasal devoicing as a sound change

Before we turn our discussion to PND, let us look into post-nasal voicing, (T > D / N_), the exact inverse process to PND. Post-nasal voicing (PNV) is well attested and phonetically well motivated, both as a synchronic process (cf. Pater 1999) and as a sound change. Locke (1983) identifies 15 languages, out of a sample of 197, that exhibit PNV as a synchronic process.6 The high number of languages with PNV is not surprising, as voicing of stops in post-nasal position is phonetically well motivated.

The phonetics of PNV is thoroughly investigated in Hayes and Stivers (2000). Building on previous work by Ohala and Ohala (1991) and others, the authors identify two phonetic factors that render stops in post-nasal position prone to voicing: nasal leak and “compression/rarification” by the velum (Hayes and Stivers 2000). It has long been known that coarticulation occurs in the transition from nasal to oral stops: the velum must rise from a low position to a high position, at which point it closes the nasal cavity. Just before full closure is reached, nasal leakage is present “and voicing is facilitated” (Hayes and Stivers 2000). Moreover, when the velum rises from a high position to complete closure, it increases the volume of the oral cavity, which again favors voicing (Hayes and Stivers 2000).

Not only is post-nasal voicing phonetically motivated, it is also universally present as a passive phonetic tendency: that is to say, phonetic voicing is found even in languages without phonological PNV, such as English. Hayes and Stivers (2000) show that speakers produce more passive phonetic voicing on voiceless stops in post-nasal position than elsewhere. Speakers produce “significantly more closure voicing” in words like *tampa* that in words like *tarpa*. PNV thus meets all the criteria for being a universal phonetic tendency. In this respect, PNV is very similar to final devoicing: both processes are well attested, have clear phonetic precursors and motivation, and operate passively even in languages without corresponding phonological processes.7

PNV is commonly attested not only as phonological and phonetic process, but also as a sound change. Kümmel (2007, 53ff.) lists at least 32 languages in which PNV operated as a sound change. This is, of course, expected if we assume that sound change strictly follows phonetic naturalness.

---

6 As reported in Hayes and Stivers (2000).
7 Some speakers show an increased VOT in post-nasal voiceless stops in *tampa* as opposed to *tarpa* (Hayes and Stivers 2000), reflecting another common change reported in many Bantu languages: aspiration of post-nasal voiceless stops (Hayes and Stivers 2000). Nevertheless, there is still significantly more voicing into closure in these stops (for all speakers).
While voiceless stops tend to be universally voiced in post-nasal position — a phonetically well-motivated and natural process — the opposite process, devoicing of voiced stops in post-nasal position, is unnatural and universally dispreferred. Several interesting factors come into play here. First, the PND process operates against a universal phonetic tendency. Second, none of the languages that show PND have it as a passive phonetic process. Third, for reasons discussed above (nasal leakage and increased volume of oral cavity), the post-nasal environment is antagonistic to devoicing. Voiceless stops are especially dispreferred in post-nasal position because the velum does not close completely; as a result, air leakage occurs into a portion of the following stop closure, prohibiting the “air pressure buildup” necessary to articulate a voiceless stop (for a discussion on phonetics, see also Coetzee and Pretorius 2010).

Despite its unnatural status, the existence of PND as a sound change has been reported in seven languages and dialects from four language families: Yaghnobi (Xromov 1972, 1987), Tswana and Shekgalagari (Solé et al. 2010), the south Italian dialects Sicilian and Calabrian (Rohlfs 1949), and the Austronesian languages Murik, Buginese, and Land Dayak (Blust 1974, Blust 2013). Additionally, PND has recently been discovered in Konyagi (Merrill 2016). PND has also been confirmed as a synchronic phonological process in Tswana and Shekgalagari (Hyman 2001, Coetzee et al. 2007, Coetzee and Pretorius 2010, Gouskova et al. 2011) and possibly in Buginese (Sirk 1983). In Shekgalagari, for example, voiced stops are devoiced synchronically when preceded by a nasal. Consider the following examples:

(6) Shekgalagari (from Solé et al. 2010)

\[
/χ0-m-bó̞n-á/ \rightarrow [χompáná] \quad \text{‘to see me’} \\
/χ0-m-ðʊζ-a/ \rightarrow [χontʊζa] \quad \text{‘to annoint me’}
\]

In the remainder of this section, I will assess the seven putative instances of PND as a sound change. If PND can indeed be shown to have occurred diachronically, this finding will indicate that sound change can operate even against the universal phonetic tendency and is therefore unconstrained.

Various explanations have been proposed to account for isolated cases of PND, but no account has yet attempted to examine this phenomenon from a cross-linguistic perspective. I will show below that, if we take all seven attested cases together, a common feature emerges: each language with PND presents evidence, direct or indirect, for a stage with complementary distribution between voiced stops (in post-nasal position) and fricatives (elsewhere). Building on this evidence, I argue against analyses that invoke PND as a self-contained sound change. Instead, I propose that each of the languages listed above underwent a set of either two or three well-motivated and well-attested sound changes that gave rise to apparent, but not actual, PND.

The aims of this section are twofold: first, I show that each instance of sound change is constrained to follow phonetic naturalness, without exception. Second, I provide a model for explaining similar cases of seemingly unnatural sound changes in the future. The success of this model in explaining cases of unnatural sound change is supported not only by traditional PND data but also by novel examples from Konyagi.
2.1 The data

2.1.1 Yaghnobi

PND was first proposed for Yaghnobi by Xromov (1972). Yaghnobi is an Iranian language, spoken by approximately 13,500 speakers in five different areas of Tajikistan (Paul et al. 2010, 4). It is the only living descendant of Sogdian, an Eastern Iranian language that was spoken around the fourth century CE. Xromov observes that NT sequences in Yaghnobi correspond to ND sequences in ancestral Sogdian; on the basis of this observation, he posits a sound change ND > NT in the development from Sogdian to Yaghnobi. The following table lists cognates from Yaghnobi and Sogdian that confirm this correspondence⁸ (from Xromov 1972: 128).⁹

<table>
<thead>
<tr>
<th>Sogdian</th>
<th>Yaghnobi</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ɣandum</td>
<td>ɣantum</td>
<td>‘wheat’</td>
</tr>
<tr>
<td>ʔʃkamb</td>
<td>ʃikampa</td>
<td>‘stomach’</td>
</tr>
<tr>
<td>sang</td>
<td>sank(a)</td>
<td>‘stone’</td>
</tr>
<tr>
<td>rangaːn</td>
<td>rankiːna</td>
<td>‘color’</td>
</tr>
<tr>
<td>anguːt</td>
<td>unkuːt</td>
<td>‘finger’</td>
</tr>
<tr>
<td>ʃəndər</td>
<td>ʃɨntir</td>
<td>postp.</td>
</tr>
<tr>
<td>-and</td>
<td>-ant</td>
<td>3rd pl.</td>
</tr>
</tbody>
</table>


Yaghnobi does in fact contrast NT and ND sequences, but this is likely secondary, introduced late in the language’s development through borrowings from Tajik (cf. Xromov 1972). Some examples of borrowed ND sequences from Tajik are given in Xromov (1972: 128): angíʃt ‘coal’, ʧəng ‘dust’, balánd ‘high’, lǔnda ‘round’. It is even possible to find instances of the inherited Yaghnobi vant ‘tie’ — with an unvoiced stop after a nasal — in contrast to the borrowed band ‘tie’ with a voiced variant.

2.1.2 Tswana and Shekgalagari

As already mentioned, PND has been reported a synchronic phonological process in Tswana and Shekgalagari (Hyman 2001, Solé et al. 2010), two closely related Southern Bantu languages; Tswana is spoken by approximately 4–5 million people in Botswana, Namibia, Zimbabwe, and South Africa (Coetzee and Pretorius 2010), and Shekgalagari by approximately 40,000 people in Botswana (Lewis et al. 2015).¹⁰

---

⁸ In addition to the list below, Xromov (1972) provides several place names which, according to him, also underwent the same sound change: e.g. Ankatak, Antark, Varvant, Žámpi, Ispantič, Kantuk, Tagrimpot; however, these examples are difficult to verify.

⁹ Data from older descriptions has been adjusted throughout this paper, as accurately as the descriptions allow, to fit IPA conventions.

¹⁰ Some estimate that the actual number of speakers is higher, since Shekgalagari may be spoken outside Botswana as well.
The table below shows that voiced stops become voiceless when preceded by a nasal (data from Solé et al. 2010). Voiceless stops remain unchanged both post-nasally and elsewhere. Solé (2012) assumes that this alternation is a consequence of a sound change that devoiced post-nasal voiced stops.

<table>
<thead>
<tr>
<th>No N-prefix</th>
<th>N-prefix</th>
</tr>
</thead>
<tbody>
<tr>
<td>χʊ-pak-a</td>
<td>χʊ-m-pak-a</td>
</tr>
<tr>
<td>χʊ-tot-a</td>
<td>χʊ-n-tot-a</td>
</tr>
<tr>
<td>χʊ-cób-á</td>
<td>χʊ-ŋ-cób-á</td>
</tr>
<tr>
<td>χʊ-kɛl-a</td>
<td>χʊ-ŋ-kɛl-a</td>
</tr>
<tr>
<td>χʊ-bón-á</td>
<td>χʊ-m-pón-á</td>
</tr>
<tr>
<td>χʊ-dɔz-a</td>
<td>χʊ-n-tɔz-a</td>
</tr>
<tr>
<td>χʊ-jís-a</td>
<td>χʊ-ŋ-cis-a</td>
</tr>
<tr>
<td>χʊ-at-a</td>
<td>χʊ-ŋ-kat-a</td>
</tr>
</tbody>
</table>

Several peculiarities need to be noted with respect to Tswana and Shekgalagari (cf. Solé et al. 2010). First, /ɡ/ never surfaces as a voiced stop: while it is devoiced to [k] post-nasally, it gets deleted elsewhere, e.g. [χʊ-at-a] for /χʊ-gat-a/ (cf. [χʊ-ŋ-kat-a]). Second, voiced stops in nasal clusters of secondary origin (after syncope) do not undergo devoicing, but remain voiced in Shekgalagari and undergo assimilation in Tswana.

Sh. /χʊ-m-bón-á/ → [χʊmpɔná]
Sh. /χʊ-mo-bón-á/ → [χʊmbɔná]
Ts. /χʊ-mo-bôn-á/ → [χʊmbɔná]

### 2.1.3 South Italian Dialects

Sicilian and Calabrian are dialects of Italian spoken in the corresponding regions of Italy by approximately 4.7 million speakers (Lewis et al. 2015). PND has been reported for these dialects in Rohlfs (1949, 424f.). The peculiarity about South Italian PND is that the sound change targets only the voiced affricate /ʤ/, which is devoiced to [ʧ] after the nasal n (ʤ >ʧ/N ; Rohlfs 1949, 424f.; Kümmel 2007, 376); regular voiced stops are not reported to be devoiced in the post-nasal position. The table below illustrates PND in Sicilian and Calabrian.

<table>
<thead>
<tr>
<th>S.-Ital. dial.</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>anʧɪlu</td>
<td>andʒelo</td>
</tr>
<tr>
<td>pinʧiri</td>
<td>pindʒere</td>
</tr>
<tr>
<td>chainʧiri</td>
<td>plandʒere</td>
</tr>
<tr>
<td>finʧiri</td>
<td>findʒere</td>
</tr>
<tr>
<td>tinʧiri</td>
<td>tindʒere</td>
</tr>
</tbody>
</table>
2.1.4 Buginese, Murik, and the Bengoh dialect of Land Dayak

PND has been reported in three Austronesian languages: Buginese, Murik, and the Bengoh dialect of Land Dayak. PND in the latter is simply mentioned without accompanying data (Rensch et al. 2006:69; Blust 2013); I therefore leave Land Dayak out of discussion that follows.\textsuperscript{11} Buginese is spoken by approximately 5 million people in Sulawesi, an island of Indonesia; Murik is spoken in Sarawak, in Malaysia and Brunei, by approximately 1,000 speakers. These three Austronesian languages are not particularly closely related, so we cannot attribute PND to developments in a common ancestor; it is likely that PND developed independently in all three branches.

Apparent PND in Buginese is represented by the following table showing the development of Proto-Malayo-Polynesian voiced stops (data from Blust 2013). Labial stops appear devoiced after nasals, but surface as [w] initially and word-externally (with a sporadic reflex [b] in initial position). The dental stop /d/ is not implicated in PND; Pre-Buginese *d develops to /r/ in all positions, which does not undergo devoicing, e.g. *dindin > renrin. Word-initially, however, *d is sporadically preserved as a voiced stop. The voiced fricative *z gets occluded to a voiced palatal stop initially, probably via rhotacism to *r intervocally. Post-nasally, *z it is devoiced to [c], as expected. Word-finally, all stops converge as [ʔ].

<table>
<thead>
<tr>
<th></th>
<th>#_</th>
<th>V_V</th>
<th>N_</th>
</tr>
</thead>
<tbody>
<tr>
<td>*b</td>
<td>b/w</td>
<td>w</td>
<td>p</td>
</tr>
<tr>
<td>*d</td>
<td>d/r/l</td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>*g</td>
<td>g</td>
<td>g</td>
<td>k</td>
</tr>
<tr>
<td>*z</td>
<td>j</td>
<td>r</td>
<td>c</td>
</tr>
</tbody>
</table>

The data cited as evidence of PND as a sound change in Buginese are as follows, taken from (Blust 2005, 2013):

<table>
<thead>
<tr>
<th>Proto-SS</th>
<th>Buginese</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>*bemba</td>
<td>bempa</td>
<td>‘water jar’</td>
</tr>
<tr>
<td>*lambuk</td>
<td>lampuʔ</td>
<td>‘pound rice’</td>
</tr>
<tr>
<td>*limboŋ</td>
<td>lempoŋ</td>
<td>‘deep water’</td>
</tr>
<tr>
<td>*rambu</td>
<td>rampu</td>
<td>‘fringe’</td>
</tr>
<tr>
<td>*rumbia</td>
<td>rumpia</td>
<td>‘sago palm’</td>
</tr>
<tr>
<td>*tambin</td>
<td>tampin</td>
<td>‘addition to a house’</td>
</tr>
<tr>
<td>*barumbun</td>
<td>warumpuŋ</td>
<td>‘a color pattern’</td>
</tr>
<tr>
<td>*bumbun</td>
<td>wumpuŋ</td>
<td>‘heap up’</td>
</tr>
<tr>
<td>*gengem</td>
<td>genkeŋ</td>
<td>‘hold in the hand’</td>
</tr>
<tr>
<td>*tungal</td>
<td>tuŋkeʔ</td>
<td>‘each, single’</td>
</tr>
<tr>
<td>*angap</td>
<td>ankaʔ</td>
<td>‘price’</td>
</tr>
<tr>
<td>*anŋap</td>
<td>ancaʔ</td>
<td>‘offerings to spirits’</td>
</tr>
<tr>
<td>*jaŋji</td>
<td>janci</td>
<td>‘to promise’</td>
</tr>
</tbody>
</table>

\textsuperscript{11} Land Dayak is not included in the count: the eight languages with PND exclude Land Dayak.
PND is reported not only as a sound change in Buginese, but also as a synchronic alternation in derivational morphology. Sirk (1983:35-37) shows that sequences of N + D yield NT (except for dentals), while the sequence N + [w] yields [mp]. It is also argued that the only permissible non-geminate clusters in Buginese are NT (to the exclusion of ND) (Sirk 1983:35-37). Examples from Sirk (1983:35-37) are provided in the table below.

<table>
<thead>
<tr>
<th>wərə</th>
<th>‘heavy’</th>
<th>sim-pərə</th>
<th>‘just as heavy’</th>
</tr>
</thead>
<tbody>
<tr>
<td>bone</td>
<td>pers. name</td>
<td>arum-pone</td>
<td>‘prince Bone’</td>
</tr>
<tr>
<td>gora</td>
<td>‘shouts’</td>
<td>səməŋ-kora</td>
<td>‘loud shous’</td>
</tr>
<tr>
<td>jaiʔ</td>
<td>root</td>
<td>məŋ-caiʔ</td>
<td>‘to sew’</td>
</tr>
</tbody>
</table>

In Murik, both labials and alveolars undergo devoicing in post-nasal position; velars are also sporadically devoiced elsewhere. Voiced dentals appear as [l] word-initially and [r] word-internally. PMP *z develops to the voiced palatal affricate [ɟʝ] initially, [s] word-internally, and to the voiceless palatal affricate [çç] post-nasally. The table below illustrates the development of stops from Proto-Kayan-Murik (as reconstructed in Blust 2005).

<table>
<thead>
<tr>
<th>#_</th>
<th>V_V</th>
<th>N_</th>
</tr>
</thead>
<tbody>
<tr>
<td>*b</td>
<td>b</td>
<td>p</td>
</tr>
<tr>
<td>*d</td>
<td>l</td>
<td>r</td>
</tr>
<tr>
<td>*g</td>
<td>g/k</td>
<td>g/k</td>
</tr>
<tr>
<td>*z</td>
<td>j/j</td>
<td>s</td>
</tr>
</tbody>
</table>

Similarly, PND is confirmed for Murik by the following examples (data from Blust 2005: 259f.; Blust 2013: 668):

<table>
<thead>
<tr>
<th>Proto-KM</th>
<th>Murik</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>*kelembit</td>
<td>kələmpit</td>
<td>‘shield’</td>
</tr>
<tr>
<td>*bumbuŋ</td>
<td>umpuŋ</td>
<td>‘ridge of a roof’</td>
</tr>
<tr>
<td>*lindem</td>
<td>lintəm</td>
<td>‘dark’</td>
</tr>
<tr>
<td>*-inda</td>
<td>t-inta</td>
<td>‘beneath, below’</td>
</tr>
<tr>
<td>*mandaŋ</td>
<td>mantaŋ</td>
<td>‘to fly’</td>
</tr>
<tr>
<td>*tundek</td>
<td>tuntuk</td>
<td>‘beak of a bird’</td>
</tr>
<tr>
<td>*lindiŋ</td>
<td>lintiŋ</td>
<td>‘wall of a house’</td>
</tr>
<tr>
<td>*undik</td>
<td>untik</td>
<td>‘upper course of a river’</td>
</tr>
<tr>
<td>*tandab</td>
<td>tantap</td>
<td>‘catch’</td>
</tr>
<tr>
<td>*andeŋ</td>
<td>antəŋ</td>
<td>‘deaf’</td>
</tr>
<tr>
<td>*pindan</td>
<td>pintaŋ</td>
<td>‘blossom’</td>
</tr>
<tr>
<td>*pendan</td>
<td>pəntan</td>
<td>‘small fruit bat’</td>
</tr>
</tbody>
</table>
The data presented in this subsection seem to suggest, at first glance, that PND operated as a single sound change in the development of all seven languages. However, I will demonstrate below that these data are misleading. A thorough investigation reveals a common pattern of complementary distribution in all seven cases, strongly suggesting that a combination of natural sound changes operated in place of a single PND. In section 2.2, I discuss the explanations of PND that have been proposed in the literature so far, and introduce some further difficulties faced by analyses invoking PND as a single sound change.

2.2 Explanations of PND

Several accounts in the literature understand PND as a single sound change; explanations for such an analysis run the gamut from appeals to sociolinguistic factors (Xromov 1972) to arguments that PND is actually a phonetically plausible or even natural process (Solé et al. 2010, Solé 2012). Three problems arise with such accounts: first, they all struggle to explain why a sound change should operate against the strong phonetic tendency to maintain voiced stops in post-nasal position; second, they each examine and account for only a single instance of PND, examined in isolation from relevant cross-linguistic data; third, most of them rely on the reconstruction of hypothetical, unattested dialects for which there is no comparative evidence.

Xromov (1972) invokes socio-linguistic factors to explain PND in Yaghnobi. He postulates the existence of two unattested dialects in pre-Yaghnobi, the first of which (Dialect 1) voiced all post-nasal stops, and the second of which (Dialect 2) retained the contrast for voice in post-nasal position. To arrive at PND, Xromov suggests that speakers of Dialect 2 generalized voiceless stops in post-nasal position based on hypercorrection to Dialect 1. The stages of this process are represented schematically in the following table:

<table>
<thead>
<tr>
<th>Dialect 1</th>
<th>Dialect 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>ND</td>
</tr>
<tr>
<td>Stage 2</td>
<td>ND</td>
</tr>
</tbody>
</table>

The major drawback of this explanation is its total lack of evidentiary basis; there are no data providing support for the reconstruction of Xromov’s two hypothetical dialects. Furthermore, if such dialects had existed, it remains unclear on Xromov’s account why the postulated hypercorrection should have occurred in the phonetically disfavored direction of PND.

Blust (2005, 2013) offers three possible explanations for the emergence of PND in Buginese, Murik, and Land Dayak. First, he notes that, much as PNV can be understood as an assimilation of stops to a voiced environment, PND can be explained as dissimilation. This assumption,
however, lacks explanatory power: it simply restates that PND is the opposite process from PNV. Blust (2013: 668) himself notes that “this does little to explain why a change of this type would occur.” Moreover, as Hayes and Stivers (2000) show, PNV is not merely an assimilation, but rather a phonetic tendency.

Blust’s second explanation for Austronesian PND postulates that the three languages in question first underwent PNV: voiceless stops became voiced in post-nasal position, thus eliminating NT sequences. According to Blust (2013), after the shift to PNV, “voice was free to vary” post-nasally, and the “voiceless variant of postnasal obstruents prevailed over time.” There are two major issues with this approach. First, it is difficult to explain why a voiceless variant would prevail in an environment that strongly favors voicing. Second, it not parsimonious to assume the independent occurrence of PNV three times without any comparative evidence.

A third explanation offered in Blust (2005) invokes dissimilation by hypercorrection. Blust notes that NT sequences in Buginese and Murik develop either to T or TT. This means that, at a certain point, NT sequences were absent from the language and only voiced stops surfaced after nasals (ND). At this point, according to Blust, speakers “may have assumed that prenasalized obstruents had acquired voicing by assimilation” and then “undid” that assumed voicing. This account faces three major difficulties. As already pointed out by Blust (2005), it is unclear what would “prompt speakers to assume that voicing assimilation had taken place in earlier clusters” of ND. Second, even if they had made this assumption, the speakers would still have to apply dissimilation in a phonetically unnatural direction. Finally, this approach lacks broader explanatory power, since it cannot be extended to cases of seeming PND in other languages, where the sound change NT > TT, T is not attested.

Some analyses have attempted to account for PND by motivating the process phonetically. Solé et al. (2010) and Solé (2012) specifically identify PND as a “historical process,” meaning that they assume PND operated as a single instance of sound change. Moreover, these authors claim that PND is not necessarily an unnatural process and may in fact have a phonetic explanation. The main evidence for this claim comes from Shekgalagari, which is assumed to feature “early velic rising” in NT sequences. This process is supposed to follow from the fact that (i) speakers do not show any passive voicing in the NT sequences in Shekgalagari and (ii) underlying nasal–fricative sequences /nz/ yield a nasal affricate [nts]. This process of early velic rising, which is argued to account for both these observations, would also have caused a “long stop closure” in ND sequences. Because voicing is difficult to maintain, especially during longer closure, the result would be devoicing of the stop (Solé et al. 2010: 612).

This explanation has two major drawbacks. First, secondary ND sequences surface as NN in Tswana and ND in Shekgalagari. The following two examples illustrate this distribution (data from Solé et al. 2010):

Sh. /χʊ-m-bɔn-á/ → [χʊmpɔná]
Sh. /χʊ-mo-bɔn-á/ → [χʊmbɔná]
Ts. /χʊ-mo-bɔn-á/ → [χʊmmbɔná]
If early velic rising in Shekgalagari were indeed a phonetic process, we should expect to see devoicing in secondary ND sequences as well. The fact that the stops in secondary ND sequences surface as voiced speaks strongly against the proposal in Solé et al. (2010) and Solé (2012). Of course, one could assume that early velic rising operated prior to the period during which secondary ND sequences arose. However, there’s a major flaw in this assumption: Solé et al. (2010) provide evidence for early velic rising from synchronic phonetic data. If we postulate that early velic rising is responsible for PND as a synchronic phonetic process, we should expect secondary sequences to undergo devoicing as well. Conversely, if we posit that early velic rising should have been completed by the time secondary ND sequences were introduced, we should not expect to find continuing evidence for this process in the current phonetic data.

Second, the explanation proffered in Solé et al. (2010) neglects dialectal data from Tswana that point against the early-velic-rising proposal. Several Tswana dialects show unconditioned devoicing of voiced stops that is most likely connected to the devoicing seen in Shekgalagari. For these cases, appeal to early velic rising in NT sequences is of no value. Likewise, such an explanation fails to account for other, cross-linguistic cases of seeming PND where no traces of early velic rising can be found. If PND were indeed motivated by early velic rising, we should expect to find evidence that early velic rising is similarly responsible for reported cases of PND outside Tswana and Shekgalagari. The alternative argument pursued by this paper — that all cases of PND result from a combination of sound changes — has more explanatory power and captures the data better than the proposal in Solé et al. (2010).

Finally, Dickens (1984) and Hyman (2001) propose an explanation for PND in Tswana that assumes a set of three non-PND sound changes that conspire to produce apparent PND: fricativization, devoicing of stops, and occlusion of fricatives. As I will argue in the remainder of this paper, this is in fact the correct explanation, and an essentially similar historical scenario played out in all eight cases of reported PND described above. Unfortunately, at the time of Dickens’ and Hyman’s work, no historical parallels existed in the literature that would support their explanation, which led other authors to propose alternative accounts of the data. Admittedly, in the absence of typological parallels, one might judge an explanation that operates with a single (albeit unnatural) sound change more parsimonious and justified than an explanation that requires three separate sound changes. By bringing numerous cases of PND from disparate language families together, this paper dispels this concern and validates the three-sound-change analysis on typological grounds.

2.3 Combination of sound changes

A closer look into the collected data from 2.1 reveals an important generalization: for all seven cases, either direct evidence or clear indirect evidence can be found that, at some stage of development, voiced stops surfaced as fricatives except in post-nasal position. Below, I show that, during that stage of complementary distribution, an unconditioned devoicing of voiced stops occurred in all seven languages. Because voiced stops surfaced only in post-nasal position, seeming PND arose. The explanation I pursue, which invokes a set of two or three (natural) sound changes, accurately captures the data for all seven cases of apparent PND. As part of this analysis, I provide new direct diachronic evidence that confirms the existence of a stage with complementary distribution in the development of Yaghnobi. I also show that unconditioned
devoicing after a stage with complementary distribution is not limited to stops, but can operate on other segments as well (e.g. devoicing of affricates in South Italian dialects). This account provides a model for explaining similar cases of “unnatural” sound changes in the future: I show that the proposed explanation also holds for the recently discovered Konyagi.

The historical evidence that complementary distribution underlies PND is best preserved in Sogdian and Yaghnobi, its direct descendant. Together, these two languages directly preserve all three stages of the PND development process that I propose operates in all attested cases. The complementary distribution of voiced stops (post-nasally) and fricatives (elsewhere) is attested as a synchronic stage in Sogdian. The limitation of voiced stops to post-nasal position in Sogdian is robustly confirmed by the writing system, leaving no doubt as to its existence (cf. Sims-Williams 1987:178). It is equally clear that the Sogdian pattern developed through a sound change D > Z / [-nas]_ in an earlier stage of the language. In Avestan (which represents an earlier stage of Sogdian), voiced stops correspond to Sogdian voiced fricatives except post-nasally, e.g. Avestan dasa vs. Sogdian ðasa ‘ten’, Avestan gari- vs. Sogdian yari ‘mountain’ (Kümmel 2006), Avestan asonga- vs. Sogdian sang ‘stone’ (Bartholomae 1961:210). The evidence from Sogdian thus considerably strengthen the claim that PND proceeds through an intermediate stage with complementary distribution.

Given the complementary distribution in Sogdian, I postulate that, on the way to Yaghnobi, only one additional sound change operated: unconditioned devoicing of voiced stops (D > T). Because voiced stops surfaced only after nasals, this combination of sound changes resulted in apparent PND.

The development is summarized in the following table:

<table>
<thead>
<tr>
<th>Proto-Iranian</th>
<th>*band</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sogdian</td>
<td>D &gt; Z / [-nas]_ βand</td>
</tr>
<tr>
<td>Yaghnobi</td>
<td>D &gt; T vant</td>
</tr>
</tbody>
</table>

Sogdian thus provides indirect historical evidence showing that the apparent case of PND in Yaghnobi is a side effect of two natural and well-attested sound changes: (i) fricativization of voiced stops except in post-nasal position (D > Z / [-nas]_), and (ii) unconditioned devoicing of voiced stops (D > T).

Note that PND is not a completely systematic phonological alternation in modern-day Yaghnobi: the original voiced labial and velar stops still surface as voiced fricatives in the “elsewhere” position (e.g. vant ‘tie’ from *band). Nevertheless, the voiced alveolar fricative (ð) gets occluded in Yaghnobi and surfaces as a voiced stop [d] (Xromov 1972: 123). Apparent PND in Yaghnobi thus holds only for the dental series of stops, because only this series of stops underwent a sound change that turned the original voiced fricatives “back” to stops (Z > D). (7) illustrates the three sound changes that operated on the alveolar series of stops to produce synchronous PND.
Modern-day Yaghnobi thus offers an intriguing snapshot of a diachronic development in which voiced stops are in the process of being occluded, but only in the alveolar series. Yaghnobi thus attests another, third, stage in the development: occlusion of fricatives to stops in some, but not all, places of articulation (Murik and Buginese, too, show selective occlusion, while in Tswana and Shekgalagari, occlusion occurs for all series of stops). Taken together, therefore, Sogdian and Yaghnobi provide direct diachronic evidence for all three stages of the development process that (as I will argue in detail below) yielded PND in all seven attested cases.

Yaghnobi and Sogdian bear yet more crucial evidence that the proposal above is correct. Novák (To appear) reports that Sogdian sequences of voiced fricative plus voiced stop appear devoiced in Yaghnobi. The table below shows cognates of consonant clusters in Sogdian and Yaghnobi that have undergone devoicing.

<table>
<thead>
<tr>
<th>Sogdian</th>
<th>Yaghnobi</th>
</tr>
</thead>
<tbody>
<tr>
<td>γd</td>
<td>xt</td>
</tr>
<tr>
<td>βd</td>
<td>ft</td>
</tr>
<tr>
<td>zd</td>
<td>st</td>
</tr>
</tbody>
</table>

This process of devoicing in clusters automatically follows from our proposal and confirms that the proposal is correct: voiced stops after a fricative did not undergo fricativization, but remained stops just like their post-nasal counterparts. Lack of fricativization in this environment is motivated by a preference to avoid clusters of voiced fricatives (since such clusters are typologically and articulatorily dispreferred). We have already seen that, in Yaghnobi, a sound change regularly devoiced all voiced stops (D > T), which occurred only after nasals and, as seen in the table above, after voiced fricatives (recall that this synchronic stage is attested in Sogdian). Unconditioned devoicing targeted stops in both environments, straightforwardly explaining why Sogdian sequences of voiced fricative + stop surface as voiceless sequences of a fricative and a stop in Yaghnobi. After voiced stops underwent unconditioned devoicing, the preceding fricative assimilated in voicing too, producing an ST sequence from an original ZD. The development is illustrated in the table below.

<table>
<thead>
<tr>
<th>Sogdian</th>
<th>ZD</th>
<th>Yaghnobi assimilation</th>
</tr>
</thead>
<tbody>
<tr>
<td>D &gt; T</td>
<td>ZT</td>
<td>γt</td>
</tr>
<tr>
<td></td>
<td>ST</td>
<td>xt</td>
</tr>
</tbody>
</table>

In the other seven languages with PND, clusters are usually not allowed (or they became simplified before the emergence of PND), so we do not see devoicing anywhere other than in
post-nasal position. Devoicing in clusters in Yaghnobi offers a crucial piece of evidence that the devoicing of voiced stops that occurs in the development of PND is unconditioned; if it were not, we would not be able to explain why these clusters devoice, as there are no other clear phonetic precursors for such a development. The series of facts revealed here indicate that apparent PND in the development of Yaghnobi actually emerged from a confluence of three sound changes and did not operate as a single sound change. I now turn to Tswana and Shekgalagari, which provide dialectal evidence for the same conclusion.

If Sogdian offered clear indications that a combination of sound changes can produce apparent PND on the historical front, Tswana brings crucial evidence for this proposal on the dialectal front. There are at least three different systems of stops in the micro-dialects of Tswana. In one set of speakers, voiced stops get devoiced in all environments: no voiced stops are allowed in the system. Speakers of this system have been labelled “devoicers” (Coetzee et al. 2007). Another set of speakers changes voiced stops into fricatives in all positions but post-nasally (these speakers are called “leniters”). A third set of speakers use the so-called PND system: for these speakers, voiced stops surface as voiceless only post-nasally. The three systems of Tswana are represented below (from Zsiga et al. 2006, Solé et al. 2010).

<table>
<thead>
<tr>
<th></th>
<th>*#ba</th>
<th>*aba</th>
<th>*mba</th>
</tr>
</thead>
<tbody>
<tr>
<td>devoicers</td>
<td>#pa</td>
<td>apa</td>
<td>mpa</td>
</tr>
<tr>
<td>leniters</td>
<td>#βa</td>
<td>aβa</td>
<td>mba</td>
</tr>
<tr>
<td>PND</td>
<td>#ba</td>
<td>aba</td>
<td>mpa</td>
</tr>
</tbody>
</table>

The PND system arises precisely through the combination of two other (devoicing and leniting) systems: leniters take on fricativization except after nasals (D > Z / [-nas]_), while devoicers undergo unconditioned devoicing (D > T). Following Dickens (1984), Hyman (2001) shows that post-nasal devoicers undergo both sound changes. Because voiced stops surfaced only after nasals, the result is apparent PND. This pattern is obscured, however, by an additional change in the dialect with PND: unconditioned occlusion of fricatives (Z > D). This change crucially blurs the initial complementary distribution of consonants, with the result that the synchronic alternation becomes PND: voiced stops surface as voiceless only after nasals. Recall that this final sound change also occurred in Yaghnobi, but only for the alveolar series of stops.

The pattern of development of Tswana voiceless stops also speaks in favor of the proposed explanation. As Hyman (2001) points out, voiceless stops underwent fricativization along with voiced stops. The table below shows the development (from Hyman 2001).

<table>
<thead>
<tr>
<th></th>
<th>#_</th>
<th>V_V</th>
<th>N_</th>
</tr>
</thead>
<tbody>
<tr>
<td>*p</td>
<td>φ</td>
<td>φ h</td>
<td>p h</td>
</tr>
<tr>
<td>*t</td>
<td>ϱ</td>
<td>ϱ h</td>
<td>t h</td>
</tr>
<tr>
<td>*k</td>
<td>h, x</td>
<td>h, x</td>
<td>kx h, k h</td>
</tr>
</tbody>
</table>

The synchronic alternation is exemplified below (Hyman 2001): *p φɛŋa mp h ɛŋa ‘conquer’
These data provide yet another confirmation that complementary distribution first occurred in Tswana and Shekgalagari, in both the voiced and voiceless series of stops. The voiceless and voiced series underwent lenition except in post-nasal position, and then voiced stops underwent further changes to produce PND, whereas voiceless stops retained the complementary distribution.

So far, I have shown that the cases of seeming PND in Yaghnobi, Tswana, and Shekgalagari can be accounted for through a combination of two or three well-motivated sound changes. I now turn to a case of PND from the South Italian dialects to illustrate that such sound change combinations involving complementary distribution are not limited to stops, but can apply to other segments as well.

On the surface, the data in South Italian suggest that /dʒ/ devoices to [ʧ] only in post-nasal position. However, if we look at the development of /dʒ/ elsewhere, we observe that it gets de-occluded to [ʒ] except after nasals (e.g. fazjina > fažina). Again, we have evidence for a stage with complementary distribution. At this point, an unconditioned devoicing of voiced affricates occurs. This, too, is a well-attested and motivated sound change: voiced affricates are highly marked. Voice is difficult to maintain, especially in affricates, and one possible resolution of this markedness is to devoice the affricates.

I propose a new explanation for the emergence of apparent PND in South Italian, which postulates a two-step process: (i) de-occlusion of voiced affricates except after nasals (DZ > D /-[nas]-), followed by (ii) unconditioned devoicing of voiced affricates (DZ > TS). The table below illustrates the development. Stage 1 shows a period of complementary distribution, with deocclusion of voiced affricates except in post-nasal position. Stage 2 represents the development after unconditioned devoicing of voiced affricates.

<table>
<thead>
<tr>
<th>elsewhere</th>
<th>post-nasally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>fažina</td>
</tr>
<tr>
<td>Stage 2</td>
<td>fažina</td>
</tr>
</tbody>
</table>

Note that this set of sound changes is in principle the same as in previous cases, but here complementary distribution targets affricates instead of stops.

The emergent pattern that we have seen in all three cases of seeming PND so far can be generalized as follows: (1) a set of segments enters complementary distribution; (2) a sound change occurs that operates on the unchanged subset of those segments; (3) optionally, another sound change occurs that blurs the original complementary distribution environment. Note that, according to Kiparsky (2008), the third sound change in this series is expected to be blocked by UG, since the combination would result in an unnatural process. However, this blocking clearly does not happen: sound change that blurs the original complementary distribution and thus
produces unnatural process is limited to dentals in Yaghnobi, but operates on all stops in Tswana and Shekgalagari.

Let us now turn to the three Austronesian languages. On the surface, the data from Buginese and Murik seem to point to PND operating as a sound change. Moreover, there is no direct historical or dialectal evidence that suggests otherwise, as is the case for Yaghnobi and Tswana. If the only attested instances of PND were those found in Austronesian languages, we would likely be forced to assume the operation of a single sound change — PND. However, these languages do, at least, show clear traces of complementary distribution. Below, I will show that the set-of-three-sound-changes explanation will again better capture the data.

The major evidence against PND as a single sound change in Austronesian comes from the voiced labial stop in Buginese. Already in Proto-South-Sulawesi (from which Buginese developed), /b/ had developed to /w/ except word-initially and in post-nasal position (Mills 1975, 547). Later, the change b > w also targeted the word-initial position. Thus, at one stage in the language’s development, voiced stops surfaced only post-nasally: again, we have clear evidence for complementary distribution. From there, the development followed the trajectory described above: unconditioned devoicing of voiced stops occurred, producing apparent PND because voiced stops surfaced only post-nasally. The development is illustrated by the following example from Buginese:

<table>
<thead>
<tr>
<th>PSS</th>
<th>Pre-Buginese</th>
<th>Buginese</th>
</tr>
</thead>
<tbody>
<tr>
<td>*bumbun</td>
<td>b &gt; w /[-nas]_</td>
<td>wumbun</td>
</tr>
<tr>
<td>*wumbun</td>
<td>D &gt; T</td>
<td>wumpun</td>
</tr>
</tbody>
</table>

In Buginese, /w/ continued to surface as a non-obstuent (only two changes operated in the labial series), but the voiced velar fricative /ɣ/ underwent occlusion to [ɡ] (three sound changes operated in the velar series), thus obscuring evidence for an inter-stage with complementary distribution. Note that this is precisely the same scenario described for Yaghnobi, with the only difference being that, in Yaghnobi, it was the alveolar series of fricatives that underwent occlusion, whereas in Buginese it was the velar series. The development leading to apparent PND is illustrated in the table below.

<table>
<thead>
<tr>
<th>PSS</th>
<th>Pre-Buginese</th>
<th>Buginese</th>
</tr>
</thead>
<tbody>
<tr>
<td>*əŋɡəp *ɡiliŋ</td>
<td>D &gt; Z /[-nas]_</td>
<td>ŋəp *ɡiliŋ</td>
</tr>
<tr>
<td>*əŋɡəp *ɣiliŋ</td>
<td>D &gt; T</td>
<td>*əŋkəp *ɣiliŋ</td>
</tr>
<tr>
<td>*aŋkəp *ɣiliŋ</td>
<td>Z &gt; D</td>
<td>aŋkəp *ɡiliŋ</td>
</tr>
</tbody>
</table>

The dental series of stops in Buginese escapes PND because *d developed to [r] in all positions and, as such, became ineligible for devoicing of voiced stops. The development of *z also conforms to the proposal above: intervocally, it undergoes rhotacism to [r]; post-nasally, it occludes and devoices to [c] (according to the unconditioned devoicing of voiced stops which predictably targets the palatals); initially, it remains a fricative *j and later occludes together with [ɣ] to [j].
In Murik, complementary distribution is still attested today and can be found in the development of voiced dental stops. On the surface, Proto-Kayan-Murik *d surfaces as [l] initially, [r] intervocically, and [t] post-nasally (see table below). I propose a historical development with the following trajectory: Proto-Kayan-Murik *d lenited to [l] or [r] (probably through an inter-stage with ð) initially and inter-vocically, where they still surface as such. After a nasal, however, *d remained a stop.

<table>
<thead>
<tr>
<th></th>
<th>V_V</th>
<th>N_</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKM</td>
<td>δ</td>
<td>d</td>
</tr>
<tr>
<td>Murik</td>
<td>r</td>
<td>t</td>
</tr>
</tbody>
</table>

Again, these facts point to a stage with complementary distribution (PKM in the table above). Since voiced stops surfaced only after nasals and an unconditioned devoicing of voiced stops occurred, apparent PND is the result. I argue that the same development occurred on labials and velars too: original voiced stops fricativized to [β] and [ɣ] (except after a nasal), at which point voiced stops (surfacing only post-nasally) got devoiced (e.g. b > p, g > k). [β] and [ɣ] then underwent occlusion “back” to stops, resulting in apparent PND. The table below traces the proposed trajectory from (Pre-)Proto-Kayan Murik to Murik for a labial series of stops.

<table>
<thead>
<tr>
<th></th>
<th>V_V</th>
<th>N_</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-PKM</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>PKM</td>
<td>β</td>
<td>β</td>
</tr>
<tr>
<td>PKM</td>
<td>β</td>
<td>p</td>
</tr>
<tr>
<td>Murik</td>
<td>b</td>
<td>p</td>
</tr>
</tbody>
</table>

A peculiarity of the development in Murik is that it combines the “PND of stops” that we saw, for example, in Tswana, with the “PND of affricates” that we saw in Sicilian and Calabrian — i.e., whereas the discussed languages devoice either stops or affricates, Murik devoices both. The development of stops in this language is straightforward — it follows the usual trajectory of PND: complementary distribution, unconditioned devoicing, and then optionally occlusion to stops (as we saw above). The development of affricates is more complicated, but nevertheless revealing. PMP *z develops to *s intervocically already in Proto-Kayan-Murik (Blust 2005); this development cannot be considered part of PND, because it happens at an earlier stage. Elsewhere, *z is preserved as voiced and gets occluded to an affricate *Ϧ. Word-initially, deocclusion to *j takes place, just as in South Italian, whereas /j/ surfaces as an affricate in post-nasal position. The affricate then gets devoiced, exactly as in Sicilian and Calabrian, together with devoicing of other voiced stops. The initial fricative *j then gets occluded together with other voiced affricates in Murik to a voiced stop.\(^\text{12}\)

In sum, devoicing that operated on voiced stops operated also on voiced affricates in Murik, and this single difference caused Murik to combine both patterns: “PND of stops” and “PND of affricates.” Thus, even though Buginese and Murik offer neither dialectal nor historical evidence for complementary distribution, there is enough language-internal evidence to posit that, at one

\(^{12}\) The affricate articulation in initial position (Blust 1974) is likely secondary: there exists variation between affricate and stop articulation. It is well known that palatal stops often develop into affricates.
stage, voiced stops (and affricates) surfaced only after nasals. In fact, the original voiced labial stops surface as fricatives even today in Buginese, whereas in Murik voiced dental stop surface as lenited [r] or [l].

All sound changes assumed under the new proposal are natural: phonetically well-motivated and well-attested, with clear phonetic precursors. A survey of sound changes in Kümmel (2007) lists at least four cases in which voiced stops undergo fricativization except after nasals, plus an additional two cases in which voiced fricatives occlude to stops post-nasally. Moreover, only 2 of the 17 surveyed languages with NC sequences permit sequences of nasal + continuant phonotactically (Maddieson 1984, reported in Steriade 1991). In fact, post-nasal occlusion is a universal phonetic tendency, as the exact phonetic realization of NZ sequences always proceeds through a stage with oral closure (Steriade 1991).

Unconditioned devoicing of voiced stops is a well-motivated process too. Voicing is universally difficult to maintain during closure, and as a consequence, languages tend to devoice stops (Ohala 1983). Of 706 languages surveyed, 166 have only voiceless stops (Ruhlen 1975, reported in Ohala 1983). Unconditioned devoicing is also attested as a sound change. In Tocharian, in addition to other examples listed in Kümmel (2007), all voiced stops devoice in all positions, including the post-nasal position, e.g. Proto-Indo-European *spend- > Tocharian spänt- (Adams 1999). Devoicing that targets all contexts and places of articulation is thus, to my knowledge, the only sound changes that causes voiced stops to devoice post-nasally — and precisely this type of unconditioned devoicing is implicated in the development of PND. Apparently, the restriction against voiced stops that results in unconditioned devoicing is stronger than the tendency for post-nasal voicing; as a result, voiced stops devoice even post-nasally under this sound change.

Finally, occlusion of non-sibilant fricatives to stops is also well-attested and motivated, though perhaps not as unidirectional as the other two sound changes. Non-sibilant fricatives are typologically and articulatorily dispreferred: there are several languages without fricatives in their inventories. Moreover, fricatives require greater level of articulatory precision than other manners of articulation: compared to stops, the articulatory targets and shape of the vocal tract require greater precision for fricatives (Ladefoged and Maddieson 1996:137). Deviation from precise articulatory targets can thus lead to occlusion of fricatives to stops. Kümmel (2007) identifies at least six languages in which a sound change operated that turned voiced fricatives to voiced stops for all places of articulation, as well as several more in which occlusion is limited to a single place of articulation.

To conclude, I have argued in this section that in none of the seven reported cases of PND does this phenomenon operate as a single sound change; instead, in each case, a combination of two or three sound changes results in apparent PND. In the following section, I show that the same explanation can be used to account for the recently discovered case of PND in Konyagi as well.
2.4 Application of the model to novel data

Recently, an eighth case of PND as a sound change has been discovered in Konyagi (also known as Wamey, among others). Konyagi, a member of the Atlantic subfamily of the Niger-Congo group, is spoken by approximately 21,000 speakers in Senegal (Lewis et al. 2015). Note that Konyagi is not even part of the Bantu family, which means that it is only very distantly related to Tswana and Shekgalagari. Merrill (2016, p.c.) reconstructs a detailed picture of Konyagi’s pre-history. Notable in this reconstruction is a series of voiceless stops in post-nasal position, which correspond to voiced stops in the neighboring languages Bedik and Basari of the Tenda group (data and table from Merrill 2016, based on Ferry 1991 and Santos 1996).

<table>
<thead>
<tr>
<th>Konyagi</th>
<th>Bedik</th>
<th>Basari</th>
</tr>
</thead>
<tbody>
<tr>
<td>ë-jàmp</td>
<td>u-yàmb</td>
<td>ø-yàmb</td>
</tr>
<tr>
<td>i-ñàmp</td>
<td>ø-ñàmb</td>
<td>a- ÿmb</td>
</tr>
<tr>
<td>i-ntàw</td>
<td>gi-ndàm</td>
<td>a-ndàw</td>
</tr>
<tr>
<td>i-kònt</td>
<td>ø-hònd</td>
<td>a-xònd</td>
</tr>
<tr>
<td>ë-ncàẹnk</td>
<td>ga-ɲąñγ</td>
<td>a-ɲąŋ</td>
</tr>
<tr>
<td>ë-ñcàl</td>
<td>ø-ɲəɬ</td>
<td>a-ɲəɲ</td>
</tr>
<tr>
<td>i-ỹęẹnk</td>
<td>u-ỹąŋ</td>
<td>a-ỹąŋ</td>
</tr>
<tr>
<td>i-nkòt</td>
<td>ɛ-ŋọt</td>
<td>ɛ-ngọt</td>
</tr>
</tbody>
</table>

The model developed above for explaining unnatural phenomena (termed the “blurring cycle” below) predicts that, at some stage in the development from pre-Konyagi to today’s Konyagi, voiced stops surfaced as voiced fricatives except post-nasally. In fact, this is exactly what we find. In Proto-Tenda, all stops fricativized everywhere but in post-nasal position. This fricativization in Proto-Tenda was quite radical, targeting both voiced and voiceless stops, as well as nasal stops and geminates/ clusters. The development is illustrated below (from Merrill 2016, p.c.).

Proto-Tenda I p t c k f j x b d j g w r j y l m n ñ ñ b ɗ ʢ y
Proto-Tenda II f r j x f j x w r j y l w n ŋ y ŋ ɓ ɗ y
Basari f s j x f j x w r j y l w n ŋ y ŋ ɓ ɗ y
Konyagi f r s x f s x w l j / w l j / l w n ŋ / v r y y

Bedik and Basari indirectly preserve the Proto-Tenda state of affairs: original voiced and voiceless stops surface as fricatives except post-nasally, where they remain stops. Additionally, original Proto-Tenda fricatives get occluded to stops post-nasally: NZ sequences develop to ND, e.g. [mw] > [mb]. In Konyangi too, original voiced stops surface as fricatives (further developing to voiced sonorants in some places of articulation) except after nasals. In post-nasal position, Konyangi voiced stops remain stops. The following table summarizes the development of consonants in post-nasal position in the descendants of Proto-Tenda (from Merrill 2016, p.c.).

Proto-Tenda I p t c k f j x b d j g w r j y l m n ñ ñ b ɗ ʢ y

25
While in Bedik and Basari post-nasal voiced stops remain voiced, they devoice in Konyagi. Here too, however — like in all other cases of PND — the voiced stops that devoiced post-nasally were in fact the only voiced stops in the language. Thus, again, the sound change operating in Konyagi is in fact unconditioned devoicing of voiced stops. The apparent PND is yet again the result of a combination of sound changes.

The third sound change of the full blurring cycle, occlusion back to stops, is lacking in Konyagi. As a result, the synchronic alternation in question is not between voiceless and voiced stops, but between voiced fricatives and voiceless stops. Konyagi, however, does not lack voiced stops completely: original fricative geminates were later occluded to stops and simplified, resulting in voiced stops being reintroduced into the synchronic inventory.13

Synchronic alternation is still active in adjectives in Konyagi, depending on the prefix. Sonorant-initial adjectives occur after vowel-final prefixes, while voiced-stop-initial adjectives occur after consonant-final prefixes. Finally, after a nasal-final prefix, we get voiceless stops (data and table from Merrill 2016).

<table>
<thead>
<tr>
<th>-V</th>
<th>-C</th>
<th>-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>-wänkáák</td>
<td>-bänkáák</td>
<td>-pänkáák</td>
</tr>
<tr>
<td>-yò lëxó</td>
<td>-gò lëxó</td>
<td>-kò lëxó</td>
</tr>
<tr>
<td>-wünkëx</td>
<td>-günkëx</td>
<td>-künkëx</td>
</tr>
<tr>
<td>-lò mâxó</td>
<td>-dò mâxó</td>
<td>-tò mâxó</td>
</tr>
</tbody>
</table>

In sum, the case of Konyagi again shows that cases of apparent PND arise from a combination of sound changes, just as our model predicts. I now turn to a discussion of PND as synchronic phonological process.

3 Post-nasal devoicing as a synchronic phonological process

We saw in section 2 that a combination of natural and phonetically motivated sound changes can result in a phonetically unnatural outcome: PND. The question now arises whether such an unnatural alternation can be incorporated into the synchronic grammar as a phonological process, and if so, to what degree such a process is productive. For final voicing, it has been claimed that no grammar exists that can permit such a process as a synchronic alternation. The best example, from Lezgian, is at least problematic, and alternative approaches are possible. Moreover, voiced and voiceless stops do not neutralize completely in word-final position. Also, there are no nonce-

13 Nasal is lost before voiceless stops.
14 Another source of voiced stops in Konyagi is Proto-Tenda nasal + implosive sequences.
words studies that would test the productiveness of final voicing. The aim of this section is to illustrate that, unlike final voicing, PND is a productive phonological process. I summarize findings from Hyman (2001), Solé et al. (2010) and Coetzee and Pretorius (2010) that clearly show PND to be productive as a synchronic process in Tswana and Shekgalagari. Unlike for final voicing, voiced and voiceless stops neutralize completely in these languages in post-nasal position, are productively extended to nonce-words, and no alternative explanations are forthcoming. Tswana and Shekgalagari thus provide the most informative example for the discussion of the relationship between synchrony and diachrony and for the question of how to derive phonological typology. In Buginese, PND is also reported as a productive synchronic alternation, but the data are sparse, so I set them aside here: further research with detailed phonetic analysis and nonce words is needed before a detailed analysis of the Buginese situation can proceed. In Yaghnobi, and Murik, PND is not productive anymore. This fact is, however, also informative by itself and is discussed in section 5.

The putative status of PND as a productive synchronic phonological process has received quite some attention since Hyman’s (2001) description of PND in Tswana. While some scholars have argued against the existence of PND (Zsiga et al. 2006, cf. Gouskova et al. 2011), most studies have provided strong evidence to suggest that PND is a productive synchronic phonological process (Coetzee et al. 2007). The most important contribution in this respect is provided in Coetzee and Pretorius’ thorough (2010) phonetic and phonological study of the phenomenon, which experimentally tests the productivity of PND using nonce-words and measures the degree to which the process is applied to novel vocabulary.

The first thorough phonetic study of Tswana PND as a synchronic process is presented in Coetzee (2007). Twelve speakers were recorded, four of which had a clear PND pattern. The other six were either leniters (see above) or had voiced stops preserved in all positions. The experiment only dealt with the four speakers with PND: the sequences of voiced and voiceless stops they produced in post-nasal position (m-pV, m-bV), and post-vocalic position (re-pV, re-bV) were recorded and analyzed (Coetzee et al. 2007). Note that nasals in both Tswana and Shekgalagari are always realized as syllabic: because clusters are disallowed, nasals become syllabic in an [mp] sequence. The results are straightforward, and strongly point to the fact that voiced stops in Tswana devoice and completely merge with the voiceless series post-nasally: /m-bV/ clearly “patterns” like /m-pV/ and /re-pV/, the only difference being the closure duration (longer after vowels). In terms of voicing percentage, the two do not differ; indeed, /m-bV/ has an even smaller VOT, which make it even less voiced. Moreover, “the contrast between /m+bV/ and /re+bV/ is significant in percentage voicing and VOT” (Coetzee et al. 2007).

Very similar results are presented in Solé et al. (2010), who offer a phonetic study of one speaker of Shekgalagari. From the results, it is clear that in Shekgalagari, too, devoicing occurs and is confirmed on all phonetic levels: it occurs “categorically,” which means that voiceless and devoiced stop are phonetically completely neutralized in post-nasal position. Voiceless and devoiced (underlying voiced) stops in post-nasal position pattern together and show no statistically significant differences in any of the four measured categories: closure duration, preceding nasal duration, VOT, and vocal fold vibration during closure (Solé et al. 2010). Voiced stops, on the other hand, show statistically significant differences along all four
parameters. This means that post-nasal devoiced stops are phonetically identical to post-nasal voiceless stops and contrast with post-nasal voiced stops (from secondary ND sequences).

Solé et al. (2010) argue for a phonetic account of PND. It is, however, unnecessary and unparsimonious to maintain such a position. As is clear from the dialectal data in Tswana and from other cases of PND discussed above, PND is the result of three sound changes and there is therefore no need to motivate PND phonetically. As Coetzee and Pretorius (2010: 417) point out, all phonetic factors are “antagonistic” to voicing in post-nasal position. This fact is not the only difficulty faced by the proposal in Solé et al. (2010). Solé et al. (2010) argue that early velic rising is responsible for PND (see above). The fact that voicing into closure in Shekgalagari is no more pronounced in post-nasal than in post-vocalic position is argued by the authors to suggest early velic rising. However, if early velic rising (and consequently PND) were indeed a synchronic phonetically motivated process, we would expect devoicing to operate on secondary ND sequences as well. This does not happen: in Shekgalagari, the secondary ND sequence remains unchanged (for discussion, see 2.2 above). One of the aims of this paper is to show that attempts to phonetically motivate a process that works against a universal phonetic tendency are unnecessary.

Both phonetic studies of Tswana and Shekgalagari come to the same conclusion: that PND is a synchronic process and that neutralization for voice is complete and categorical in post-nasal position. A question that remains, however, is whether an unnatural process that results from a combination of sound changes can become incorporated into the synchronic phonology so that it is completely productive, applying freely to novel words and nonce-words.

The major contribution to the question of productivity of PND is offered in Coetzee and Pretorius (2010). Again, 12 speakers of Tswana were recorded and the experiment included both native vocabulary and nonce-words. The paper offers two major contributions. First, the authors show that PND is extended to nonce-words at the same rate as it applies in the native vocabulary in Tswana (Coetzee and Pretorius 2010: 411). This means that PND is not lexicalized, but rather it is a productive phonological process in the synchronic grammar of Tswana. Thus, we can see that unnatural alternations produced by combinations of sound changes can incorporate into synchronic grammars and become productive.

Second, the paper shows that the natural process of PNV operates passively even in cases in which PND is a phonological process: “all of our post-nasal stops, whether underlyingly voiced or voiceless, were realized with voicing during the initial part of the stop closure” (Coetzee and Pretorius 2010: 417). The authors continue to note that some speakers nevertheless have voiceless (devoiced) stops post-nasally: voicing there is only passive and does not extend through more than half of the closure (Coetzee and Pretorius 2010: 417). This system — the one containing the unnatural synchronic phonological process PND — is attested for seven of the twelve speakers. For the other five speakers, however, Coetzee and Pretorius (2010: 417) observe that they often voice the whole closure. These results suggest that these speakers have introduced a new rule into their system: the natural, phonetically motivated, and exact inverse process to PND — post-nasal voicing. This natural tendency to voice post-nasal stops is expressed to different degrees and varies from speaker to speaker. For one speaker, almost all

---

15 If we do not interpret the data in Coetzee and Pretorius (2010) as dialect mixing.
post-nasal stops were voiced, with less than 25% of stops being voiceless in post-nasal position. It is thus clear that, at least in Tswana, PNV is operating even in systems that have PND as a synchronic phonological process.

4 A Probabilistic Model of Typology

The data presented above show that in none of the eight reported examples of PND does this process operate as a single sound change. Instead, in all eight cases, a combination of either two or three natural sound changes operated historically, together giving rise to apparent PND. These findings suggest that individual sound changes are constrained: they may not operate against the universal phonetic tendency. Moreover, PND is often cited in the literature as an example of sound change that operates in the unnatural direction. By showing that all reported cases of PND actually arise through a combination of natural sound changes, we can maintain the position that any single instance of sound change has to be natural.

Three natural sound changes, however, can occur such that the result of the combination is an unnatural alternation. The aim of this section is to explain why unnatural alternations are rare or unattested despite the fact that any combination of single sound changes is possible. The standard explanation offered for typology within the channel bias approach is that “common sound patterns often reflect common instances of sound change” (Blevins 2013:485, also Greenberg 1978:75-6). In other words, the more common a sound change is, the more common the synchronic alternation that it will produce. However, by assuming only this factor, we face problems motivating the wide spectrum of pattern frequencies — in particular, why some patterns are very common, whereas others are unattested. Below, I propose a model that crucially relies on the distinction between a single instance of sound change and a combination of sound changes and adds a probabilistic dimension to the derivation of typology. I argue that this approach captures surface typology better than the proposals entertained so far: I explain why natural alternations are the most common, unmotivated alternations less common, and unnatural alternations rare or even unattested.

The idea that unusual rules are rare because they require complex history is not new. The low probability of combinations of changes has been previously relied on to account for the rarity of certain morphological processes, and some attempts have been made to use this reasoning in phonology as well (Greenberg 1978:75-6, Cathcart 2014). Blevins (2004:310) briefly mentions that the rarity of certain morphological processes (such as tense marking on pronouns in Gurnu) might be explained by the low probability of co-occurrence of the factors that led to this system. This approach is employed more thoroughly in Harris (2003, 2008), who notes: “the more changes are involved, the less likely all will happen to co-occur” and “it is an idea that is necessary to discuss, because the role of probability has not been included in previous discussions of rare phenomena.” However, the proposals often do not go much further than stating this generalization; furthermore, models that do pursue such a premise yield wrong predictions because only the overall probabilities of the sound changes are calculated (Cathcart 2014), without taking into consideration the crucial distinctions made in this paper. The subdivision of unusual rules to unnatural versus unmotivated rules, paired with the proof that the latter require at least three sound changes to arise, facilitates the development of a more elaborate, quantifiable model. In this section, I will articulate two trajectories by which unnatural
(as opposed to unmotivated) rules can arise. The first, which I have termed the **BLURRING CYCLE**; captures the development of PND. The other hypothetical trajectory is the **BLURRING CHAIN**, which also helps us explain cases of unnatural phenomena. After schematizing these two trajectories below, I provide a proof that at least three sound changes must function together in order for an unnatural process to arise. This property is then exploited to provide an elaborate model of typology within the channel bias approach.

Let us assume that all sound changes have to be natural, but that any combination of single instances of natural change is possible. Let \( A > B \) in an environment \( X \) be a universal phonetic tendency and a natural sound change:

\[
A > B / X
\]

Let the following sound changes all be instances of natural sound change and phonetic tendencies:

\[
B > C / Z \\
B > A \\
C > B
\]

The result of the combination of the sound changes in (8) is \( B > A \), whereas \( B \) stays \( B \) in the environment \( Z \). If the environment \( Z \) is precisely the complement of \( X \), we get \( B > A / X \): the exact opposite process of the natural process in (7). This is exactly the scenario attested in the eight cases of PND discussed above.

\[
B > A / X
\]

Another possible scenario in which a combination of sound changes results in the inverse of a natural process \( A > B / X \) is illustrated in (11). Again, assume that all instances of sound change in (11) are natural and follow phonetic tendencies.

\[
B > C / X \\
C > D \\
D > A
\]

---

\[16\] It should be noted that two exactly opposite sound changes can both be natural. For example, vowel height assimilation vs. dissimilation and deletion vs. epenthesis in [ns] > [nts], [nts] > [ns] (pointed out by X, p.c.) both have clear phonetic precursors and are natural sound changes. However, this does not affect our typology in any way, as natural sound changes are predicted to be common even if the pairs can operate in both directions.
In order to distinguish the two developments described here from the phenomenon of “telescoping,” I will refer to the development in (8) as the “blurring cycle” and the development in (11) as the “blurring chain.” The motivation for the term “cycle” is clear: one of the three sound changes in (9) targets the outcome of another sound change in (9) and results in the target of that other change. The term “chain” is likewise motivated: in (11), the outcomes of a sound change are targets for following sound changes. Both developments “blur” the original complementary distribution, resulting in an alternation that operates against universal phonetic tendency.¹⁷

<table>
<thead>
<tr>
<th>BLURRING CYCLE</th>
<th>BLURRING CHAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>B &gt; C / Z</td>
<td>B &gt; C / X</td>
</tr>
<tr>
<td>B &gt; A</td>
<td>C &gt; D</td>
</tr>
<tr>
<td>C &gt; B</td>
<td>D &gt; A</td>
</tr>
</tbody>
</table>

The data from Tswana and Shekgalagari confirm that the unnatural alternation B > A / X — the result of the “blurring cycle” — can be implemented as a synchronic process B → A / X. This means that speakers can implement a process into their grammar that a single instance of sound change cannot produce. From this evidence, it follows that we have at least one example of a sound change that is more constrained than UG, and thus, a piece of evidence against the radical view of the analytic bias approach, which operates with unconstrained sound change and constrained UG.

We saw that, while a single sound change is constrained to follow phonetic naturalness, a combination of sound changes appears unconstrained: any number of single instances of sound change can operate on each other, and even if such a combination results in unnatural alternation, the synchronic grammar can still incorporate it. However, if we assume that combination of sound changes is unconstrained, we still need to explain why unnatural processes (B > A / X) are rare — as in the case of PND — or even unattested — as in the case of final voicing.

We could assume analytic bias to be responsible for the rarity of such patterns. It is conceivable that, when a combination of sound changes would result in an unnatural process, the speaker’s grammar filters this combination out. Such an analysis would explain why final voicing is unattested: any combination of sound changes that would give rise to final voicing (of which there are at least five, of shown above) is ruled out in the synchronic phonological

¹⁷ Other potential trajectories of three sound changes exist for an unnatural process to arise. For example, the first scenario in Kiparsky (2006) could be schematized as:

C > B / X
B > A / X
C > B/ –X

The result here, too, is a system in which B is contrastive in both environments X and –X, but A is limited to its unnatural environment A. Such a trajectory could be termed a “blurring series,” as C > B occurs twice in different environments. However, the question remains whether such a trajectory could develop into a synchronic alternation. To my knowledge, no cases have been attested yet, which is why I omit this possibility from the main discussion.
grammar by constraints in UG, rendering such an alternation impossible. But if this were the case, why would we not expect the same for PND? Why is PND attested as the synchronic end result of a combination of sound changes, while final voicing is not? Both processes are alterations that operate in the exact opposite direction of a universal phonetic tendency — so if speakers’ grammars filter out one alternation, why can they implement the other?

It might be possible to argue that final voicing is somehow more unnatural or more constrained in UG than PND. However, there are not many compelling arguments in favor of this conclusion. Alternatively, we could argue that devoicing is a necessary strategy in the perceptual repair of some unstable contrast (Hamann and Downing 2015), for example between ND and N (as argued in Kirchner 2000). Here, too, however, problems arise. Should devoicing really be a viable repair strategy, we would expect it to be attested as such, independently of the three sound changes outlined above. There are no instances of PND as a repair strategy: all cases of PND that we have are the result of a set of sound changes. It seems suspicious to suggest that a problematic contrast will only get repaired when a set of three (unrelated) sound changes happen to operate in the pre-history of a system. Moreover, PND is not attested as a repair strategy even in cases where we should expect it. For example, many languages disallow NC₁VNC₂ sequences. A recent study in Stanton (2015) suggests that avoidance of these sequences constitutes a strategy to repair the contrast in NC₁VNC₂. The vowel in NC₁VNC₂ is universally phonetically nasalized, a process which reduces cues of the contrast between NC and N. One way to repair this contrast would be to devoice the consonant. However, the survey in Stanton (p.c.) show that NDVNC > NTVNC is not attested. Finally, in pre-Konyagi, NT and ND sequences contrast. It would be highly problematic to assume ND devoiced to NT in order to repair an ND – N contrast while simultaneously completely merging the NT ~ ND contrast. It is therefore difficult to maintain the position that PND is permitted to surface due to its role as a repair strategy.

If PND is attested as synchronic phonological process despite its violation of universal phonetic tendency, perhaps final voicing is not impossible, but simply possible and unattested. Unless new evidence is provided to show that the two processes are radically different, then, if one is attested as a fully productive phonological process, we should assume the other to be possible as well.

Let us assume that both processes are possible and that PND is simply very rare whereas final voicing is unattested. Any valid theory must still address the overall rarity of these unnatural alternations. If combination of sound changes is unconstrained, as I have argued above, we should in principle expect unnatural alternations to be just as common as natural and phonetically well-motivated processes.

To solve this problem, I propose a new model for deriving the typology of phonological processes that avoids heavy reliance on UG and analytic bias. First, it is important to note that even proponents of the analytic bias approach acknowledge that the high frequencies of some alternations arise due to channel bias. For example, Kiparsky (2006) points out that the frequent alternation k → tʃ before front vowels is not due to any particular “UG-based account,” but rather to the fact that tʃ “is the end of a chain of natural processes set in motion by the palatalization of k in that environment” (Kiparsky 2006:220). If channel bias is responsible for
explaining the typology of at least some phonological alternations, then parsimony dictates that we should attempt to explain all phonological alternations within this approach; a theory that does not introduce new theoretical apparatus is preferable. My contribution in this discussion is to show that channel bias encounters no major difficulties in explaining surface typology, contrary to what has been claimed in the literature.

Crucial to the process of deriving surface typology using the channel bias approach is the distinction between a single sound change and a combination of sound changes. Capitalizing on this distinction, we can explain the rareness of alternations such as PND and final voicing as follows: the probability of a combination of sound changes operating in such a way that they influence each other and yield an unnatural result is relatively small, and always smaller than the probability of a single sound change. Furthermore, even when a combination of sound changes does result in an unnatural alternation, the opposite phonetic tendency will begin operating against it, ultimately resulting in the likely loss of the unnatural alternation.

Another crucial aspect of this typological derivation is the fact that at least three sound changes are needed for an unnatural process to arise, whereas unmotivated processes require (at least) two sound changes and natural processes can arise through a single sound change. This paragraph provides a proof that unnatural processes require at least three sound changes to operate in combination. Let us first assume that a single instance of sound change means a change of one feature in a given context. For a natural sound change (A > B / X), A and B differ in exactly one feature (for example [±voice] in the case of PND or final voicing). How do we get the unnatural B > A / X? With one single natural sound change, it is impossible, because B > A / X is by definition unnatural. However, B > A / X is also impossible with two natural sound changes. Why? We know that A and B differ in one feature only. For a B > A / X sound change to arise, therefore, we first need B to change into something other than A (it cannot change to A directly because such sound change is unnatural). So, let B change to C, where B and C differ in one feature, but a different feature from the one that separates A and B. From this point, it is impossible to for an unnatural sound change to arise without a third sound change: C cannot develop directly to A, since the two segments differ in two features: feature F₁, which distinguishes A and B, and feature F₂, which distinguishes B and C. By definition, two sound changes are required in order to change two features; hence it follows that at least three sound changes must take place in order for an unnatural process to arise.

18 There is another advantage of this approach: it has been noted that once the sound change [k] > [kʲ] / [+front] occurs, the likelihood of further development to [cç] and [tʃ] is very high (X, p.c.). If we assume that these sound changes proceed in series, a generalization can be easily derived: the initial sound change [k] > [kʲ] / [+front] has a certain probability, which is smaller than the probability of further development to [cç] and [tʃ]. Because [kʲ] > [cç] and [cç] > [tʃ] have comparatively very high probabilities, the development to [tʃ] is high once [k] > [kʲ] / [+front] occurs.

19 If we do not assume that a sound change is necessarily a change in a single feature, but can involve changes of more than one feature simultaneously, then unnatural alternations can arise through two sound changes only: (1) B > C / X and (2) C > A / X, assuming that both (1) and (2) are natural and that C > A / X is a (natural) sound change that involves a simultaneous change of more than one feature. However, typology still follows from such an assumption: even if we admit simultaneous sound change of multiple feature in the set of possible sound changes, they are still less frequent than sound changes that target only one feature. The overall probability of an unnatural alternation will thus nevertheless be smaller than the probability of an unmotivated alternation.
This restriction is crucial for the proposed probabilistic model. Let us see how the new model works schematically. First, in a given phonological system, let the probability that a sound change \( A \) will happen be \( P(A) \). \( P(A) \) depends on several factors, one of which is necessarily the phonemic inventory of the language. For example, devoicing of stops has a probability of zero in a system without voiced stops. Calculating the probability \( P(A) \) is not a trivial task, and at the moment we do not have good estimations for probabilities of sound changes or the factors that play a role in determining those probabilities; further research on probabilistic approaches to sound change, however, should bring better tools for calculating this parameter (Bouchard-Côté et al. 2013, Gilman 2015). However, phonological typological predictions can be derived from the proposed model even without precise estimations of sound change probabilities.

I have shown above that an unmotivated alternation requires at least two sound changes and that unnatural alternation requires at least three, whereas a natural alternation can result from a single sound change. Crucially, the probability that a combination of sound changes \( A \) and \( B \) will happen must be no greater than \( P(A) \), because this probability is the product of the probabilities of each individual sound change, and therefore a subset of \( P(A) \).

\[
P(A \cap B) = P(A)P(B) \leq P(A)
\]

The probability that three sound changes will happen is even smaller: \( P(A \cap B \cap C) = P(A)P(B)P(C) \). To illustrate this point, let us assume for the sake of argument that a single instance of a natural sound change A has a probability \( P(A) = .1 \) in a given timespan \( t \) (for computing frequencies of sound changes, see Gilman 2015). In operational terms, then, sound change A has a 10% probability of occurring in a given language within a given timespan. Now consider the probability that a combination of three sound changes (A, B, and C), each with an individual probability of .1, will occur: the probability of the union of all three changes is the product of all three probabilities, equaling .001. The formula is given in (13).

\[
P(A \cap B \cap C) = .1 \times .1 \times .1 = .001
\]

The probability that a combination of sound changes will happen is thus 100 times smaller than the probability that a single sound change will happen.

Two assumptions that underlie this reasoning must be addressed here. The first is that operations of individual sound changes are independent events. This assumption is not problematic (cf. Greenberg 1978:75-6 for his proposal of “memorylessness”). It is likely that sound changes indeed operate independently, although further research is needed in this area. Indeed, it would be hard to explain how the occurrence of sound change A could influence the occurrence of sound change B, especially if sound change A does not alter phonological structure significantly and thus influence the probability of B. For the purpose of argumentation, we can make such a stipulation. A second assumption is that the probability of a combination of sound changes \( P(A \cap B) \) is only smaller than \( P(A) \) if \( A \) is a member of the set of sound changes.
that we’re comparing it too. In other words, no inherent relationship exists between \( P(A \cap B) \) and \( P(C) \). For example, let us assume \( P(C) = 0.001 \), \( P(A) = 0.2 \), and \( P(B) = 0.2 \). \( P(A \cap B) = 0.04 \) will in fact be greater than \( P(C) \). However, it is nevertheless straightforward to maintain the position that unnatural alternations are rarer than natural ones (in other words, that combinations of sound change are rarer than single sound changes). The logic is as follows: in order for the probability of a single instance of sound change \( (P(D)) \) to be smaller than a probability of a combination of three sound changes \( P(A \cap B \cap C) \), \( P(D) \) must be very low indeed, \( P(A) \), \( P(B) \), and \( P(C) \) must be (comparatively) very high. While such a situation is theoretically possible, in practice, we do not expect many natural sound changes to have extremely low probabilities of occurrence. The position and argumentation outlined above can thus be maintained: significant discrepancies in sound change frequencies are rare, and globally the necessary generalization will hold even if we allow exceptions whereby a natural sound change has extremely low probability of occurrence.

The probabilistic model thus not only predicts that unnatural alternations will be rare and natural ones common, but it also captures the continuum of increased probability between unnatural, unmotivated, and natural changes. As already mentioned above, unmotivated changes are comparatively frequent in languages and arise through “telescoping.” They do not necessarily operate against universal phonetic tendency, but they lack any obvious phonetic motivation. As was shown in Kenstowicz and Kisseberth (1977: 64), unmotivated alternations can arise from as few as two sound changes that affect each other. For unnatural processes that operate against universal phonetic tendency, on the other hand, we need at least a combination of three sound changes (via the “blurring cycle” or “blurring chain”) and only a subset of those combinations will result in an unnatural process. From this, it automatically follows that natural alternations will be the most common, because they arise from a single sound change. Unmotivated alternations are predicted to be less common than natural alternations, because they can only arise through a combination of sound change; by the same token, any combination of sound changes that affect each other will likely result in an unmotivated process. Unnatural alternations operating against universal phonetic tendency are predicted to be the most rare because at least three sound changes have to operate in combination to yield an unnatural alternation.

However, the low probability is not the only reason why unnatural processes are rare. Crucially, as soon as an unnatural process operating against universal phonetic tendency does arise and become fully and productively incorporated into the synchronic phonological grammar \( (B \rightarrow A / X) \), the inverse universal phonetic tendency \( (A > B / X) \) will begin operating against it. As a result, the probability that an unnatural alternation will survive is even further reduced by the fact that a common sound change (and universal passive phonetic tendency) operates progressively against its existence.

This erosion is precisely what we see happening in Tswana: in a system with unnatural alternation (PND), a single instance of natural and opposite sound change (PNV) is in the process of imposing its will against the unnatural alternation.

In five cases in my typological survey (South Italian dialects and Land Dayak), PND operates in the absence of the original environment of complementary distribution that produced it. In two of these cases, the PND alternation has ceased to be productive, due to the universal tendency to voice post-nasal stops militating against it or due to borrowing; in the other two (or
three, if we take Buginese into the count) cases, PND has been incorporated productively into the synchronic grammar, but sound change in the opposite and natural direction is in the process of destroying this alternation.

If, by a historical accident, Tswana and Shekgalagari had not been attested, we would continue to consider PND to be impossible. Because of Tswana and Shekgalagari, we hold PND to be a rare, but possible, synchronic process. It is quite likely that final voicing is unattested by the same historical accident. It is possible that final voicing is even rarer than PND, perhaps because final devoicing is itself a frequent and strong phonetic tendency and sound change; however, the fact that one of these processes is rare and one unattested need not be the consequence of any UG constraints. Using the model developed in this paper, we can now explain why final voicing is unattested as a synchronic phenomenon. All five scenarios for the potential development of final voicing presented in Kiparsky (2006) require three sound changes or more (as predicted): in fact, one scenario requires analogy, others require typologically unusual unusual stages (resulting in a low probability, since the needed scenario will apply to only a subset of languages), and still others do not result in synchronic alternation but in phonotactic restrictions. We saw that combinations of three sound changes that are all active in a given time frame \( t \) (for temporal limitation, see discussion below) are very rare: the fact that final voicing is not attested may well be a consequence of this rarity. Moreover, it is likely that the kinds of sound changes that target exclusively final stops — and target them in a way that would provide grounds for the blurring cycle and blurring chain — are not common, thus providing an additional explanation for the absence of final voicing.

One could argue that, even though combinations of sound changes are less frequent than any single instance of sound change, in the course of an almost unlimited timespan, sound changes ought to “stack up,” yielding multiple unnatural alternations in any given language; given that every language has a several-thousand-year history during which sound changes have occurred continuously, we should perhaps expect many more unnatural alternations than are actually attested. Consider, however, that any given sound change has a time of operation \( t \). In other words, a sound change becomes active at one point in time and ceases to operate at another point in time. This is primarily evident from the fact that, at some point in any language, certain sound changes cease to apply to novel vocabulary, loanwords, and morphological alternations. For an unnatural phonological alternation to arise, all the sound changes that play a part in this alternation must be active simultaneously. Thus, the timespan available to produce such unnatural alternations is not unlimited, but rather limited by the time \( t \) in which all single sound change that combine to yield the alternation in question are active. In probabilistic terms, we would say that language history is not a pure-birth process, but rather a birth-death process.

The model proposed here is easily quantifiable, but finer models of particular sound change probabilities are necessary for its full execution. Using the model, we can calculate the probability of any synchronic alternation based on the probabilities of the sound changes necessary for that alternation to arise. The first stage in this calculation is to identify all possible diachronic trajectories that would produce the sound change in question (let us call them T).

For a more elaborate model, we could use the Poisson stochastic process for calculating probabilities of a series of ordered independent events in a given time frame (cf. Eid 2011). Let
λ, be a rate of a sound change (in a time frame t_i), which can be calculated based on the number of occurrences per time frame we have attested in a given language. Let t be the time frame of occurrence of a given trajectory T_1 composed of three sound changes that result in an unnatural process. The probability that an unnatural process based on trajectory T_1 will arise from three sound changes in a given time frame from 0 to t equals (cf. Fussell 1976 and Eid 2011):

\[ P(T_1) = \int_0^t f_1(t) dt_1 \cdot \int_{t_1}^t f_2(t_2) dt_2 \cdot \int_{t_2}^t f_3(t_3) dt_3 \]

where

\[ f_i = \lambda_i e^{-\lambda_i t} \]

There also exists a simpler and easier-to-implement model that disregards the temporal aspect. For each trajectory, we can calculate its probability as a simple product of the probabilities of each sound change. For example, for an unnatural alternation, the trajectory with the fewest sound changes needed will include at least three sound changes. The probability P(T_1) will be a product of the probabilities of each of the sound changes divided by a number of possible orders.

\[ P(T_1) = \frac{P(A_1)P(B_1)P(C_1)}{3!} \]

Or more generally:

\[ P(T_1) = \frac{P(A_1)P(A_2)P(A_3)\ldots P(A_n)}{n!} \]

There may also be other viable trajectories for the synchronic alternation, which may require a higher number of sound changes. Regardless of the method we choose for calculating the probabilities of particular trajectories, the overall probability of a synchronic alternation P(Alt) can be calculated as the sum of the probabilities of each viable trajectory:

\[ P(Alt) = P(T_1 \cup T_2 \cup T_3 \cup \ldots \cup T_n) \]
The formula for the sum of three probabilities is:

\[
P(Alt) = P(T_1 \cup T_2 \cup T_3) = P(T_1) + P(T_2) + P(T_3) - P(T_1 \cap T_2) - P(T_2 \cap T_3) - P(T_1 \cap T_3) + P(T_1 \cap T_2 \cap T_3)
\]

Note that each alternation can arise through an indefinite number of trajectories. However, because probabilities of higher order trajectories (i.e. the ones that require higher numbers of sound changes) will be so minor, we can disregard them for the purposes of calculating probability. The formula in (19) provides grounds for calculating “historical probabilities” of phonological alternations. As probabilities for some alternations will quickly become very low, we can also calculate “historical scores” of alternations (\(P'(Alt)\)):

\[
P'(Alt) = -\log (P(Alt))
\]

This can be useful in many further respects: as we increase the precision with which we can calculate the probability of a single sound change, we can use this parameter to compare influences of channel bias and analytical bias in phonological typology. Such an analysis is, however, beyond the scope of this paper.

The argument in favor of channel bias pursued here is not meant to suggest the non-existence of learning biases, tendencies, and preferences in the learning process: the proof of existence of channel biases in phonological typology does not automatically exclude all other possible mechanisms. In fact, several studies have shown that learning bias does exist; some processes are more difficult to learn than others (see Hayes et al. 2009 and the literature therein). Moreover, a growing body of research acknowledges that both aspects play a role in typology (Moreton 2008). At the same time, it has also been shown that, given enough input, more or less anything can be learned. The present study supports this finding, as it offers several examples of unnatural alternations that have been learned as fully productive synchronic alternations. A combination of three active sound changes in a history of a language provides enough input for an unnatural alternation to be learned, and once it has been learned, it is not easy to see how learning bias might militate against such unnatural system. Obviously, more research is needed, but the model that I proposed above faces no particular difficulties on this front. A channel-bias explanation of PND is well motivated, relies on verifiable claims, and employs external (and therefore not circular) evidence based on relative frequencies of combinations of sound changes and phonetic tendencies that destroy unnatural processes. The proposed calculation of historical probabilities offers an important new framework for disambiguating the two factors.

Incidentally, this model also accounts for the phonetic groundedness of most phonological alternations as a natural outgrowth of their development: since, probabilistically speaking, most alternations will arise a single instance of sound change, they are highly likely to be natural and phonetically motivated. By contrast, the analytic bias approach faces problems accounting for for
the fact that precisely those alternations that arise through a combination of sound changes are considered unlearnable/ruled out by UG.

5 Conclusion

The main goal of this paper has been to discuss how phonological typology arises and to propose a new model for deriving that typology within the channel bias approach. I discussed two undisputedly unnatural alternations, final voicing and PND, and argued that final voicing is not the optimal example for deciding between the analytic and channel bias approach. Instead, I discussed PND and explored the implications of this process for the derivation of phonological typology. Based on data from eight different languages that exhibit PND, I presented a model that permits, on the one hand, single instances of sound change that are strictly natural and phonetically motivated and, on the other hand, combinations of sound changes that appear to be unconstrained. I showed that this model, which follows the channel bias approach, has no difficulties deriving the typology of uncommon or unattested patterns.

In the first part of the article, I addressed the status of post-nasal devoicing (PND) as a sound change. I discussed all known cases of PND and pointed to common patterns among those cases that provide evidence against analyses that invoke PND as a sound change. I argued that so far, all explanations offered for PND have been limited to particular languages (or language families), and have therefore failed to provide a good typological rationale. The range of explanations inspired by these isolated cases variously invoked sociolinguistic factors, reconstructed hypothetical dialects for which there is no comparative data, or argued that PND is a natural phonetic process. The present paper argued instead for a unified account of all eight cases, demonstrating that the explanation proposed for Tswana in Dickens (1984) and Hyman (2001) is essentially correct and pertains cross-linguistically.

Building on these data, I argued that in all eight attested cases, apparent PND is a side effect of two or three phonetically well-motivated and well-attested sound changes. The strongest argument in favor of this proposal is the fact that, in all eight cases, either direct evidence or clear traces of complementary distribution can be found, revealing a developmental stage in which voiced stops surfaced as fricatives except in post-nasal position. The best evidence for such a stage comes from Sogdian, where it is in fact historically attested. Next, I argued that, following the emergence of this complementary distribution, unconditioned devoicing of voiced stops (or voiced affricates) occurred in all eight cases. The strongest evidence for this sound change comes from dialectal data in Tswana. Some systems of modern-day Tswana undergo unconditioned devoicing of voiced stops (devoicers) and other systems undergo fricativization (leniters). The combination of the two sound changes yields PND, as has been argued previously by Dickens (1984) and Hyman (2001). Finally, a third sound change occurs optionally, blurring the original complementary distribution through the occlusion of voiced fricatives. This sound change operates in Tswana and partially in Buginese, Murik, and Yaghnobi.

The fact that PND does not operate as a single sound change, but rather is the result of a combination of sound changes, means that we do not have to posit that a single sound change can operate in an unnatural direction nor that PND is a natural and phonetically motivated process. The existence of PND also bears implications for considerations of how natural a sound change
must be. By showing that, in all cases, PND arises as the result of a combination of sound changes, we can maintain the position that a single instance of sound change is always phonetically motivated and operates in the natural direction. Additionally, I provided a model for explaining similar cases of apparent unnatural sound changes in the future: (1) a set of segments enters complementary distribution; (2) a sound change occurs that operates on the (un)changed subset of those segments; (3) optionally, another sound change occurs that blurs the original complementary distribution. I proposed the terms “blurring cycle” and “blurring chain” for the two types of possible combinations of sound changes that produce unnatural alternations. Furthermore, I showed that the model can also account for a newly discovered case of PND in Konyagi. This model has the potential to explain other unnatural developments and alternations beyond PND. Pursuing this line of research is beyond the scope of the present paper, but some unnatural alternations that will likely prove solvable using this model include intervocalic devoicing in Murik and the Berawan dialects of the Austronesian language family (Blust 1992, 2005, 2003) and voicing of stops after non-nasals in Tarma Quechua (Adelaar 1977, Nazarov 2015).

Finally, I proposed a categorical distinction between a single instance of sound change and a combination of sound changes. This distinction allowed me to build a model that correctly predicts surface typology and is grounded in external, not internal (and therefore not circular) factors. I argued that a single instance of sound change is always natural and follows universal phonetic tendencies, while a combination of sound changes may be unconstrained and can result in unmotivated or unnatural processes. I provided a proof that, while two sound changes can give rise to an unmotivated process, at least three changes are needed for an unnatural process to arise. Because a single sound change is more frequent and likely to occur, most alternations that we observe in languages will be natural. Conversely, combinations of sound changes are less probable: two sound changes in combination are less likely to occur than a single sound change; three sound changes are even less probable. As a result, unmotivated and unnatural patterns will be rare or non-existent cross-linguistically, just by virtue of the infrequency of their derivation. The proposal thus captures the continuum of decreased probability between natural, unmotivated (at least two sound changes needed), and unnatural changes (at least three sound changes needed). Furthermore, when unnatural alternations do arise, they immediately get militated against by the inverse process, which is itself a natural phonetic tendency. Thus, the rarity of an unnatural alternation may directly depend on how strong its phonetic tendencies and precursors are and on how many single instances of sound change are necessary to produce the alternation. The proposed model is also quantifiable: I provide grounds for calculating “historical probabilities” for the emergence of each phonological alternation. In future, this model has the potential to help us disambiguate between channel and analytical factors in phonological typology, but this is beyond the scope of this paper.

The paper does not claim that analytic bias does not exist: the exact relationship between the channel and analytic bias remains to be determined in future research and the paper aims to contribute to this discussion. It is likely that the two types of bias are more interrelated that they appear. The primary contribution of this paper has been to illustrate that the channel bias approach can satisfactorily derive surface phonological typology using the model I propose.
References


Gilman, Sophia. 2015. *Operationalizing the intuitive aspects of the comparative method*. Ms., Yale University. Available on Jun 6, 2015 at: https://docs.google.com/document/d/1XMmRcKe2ry6DLAJQgVkUqkbHqKfdbm_IDSdQkP4-g0/edit


Stanton, Juliet. 2015. Constraints on contrast motivate nasal cluster effects. Handout at LingLunch. May 14, 2015, MIT.


March 26, 2016