PAUL BOERSMA

THE ODDS OF ETERNAL OPTIMIZATION IN OPTIMALITY THEORY

Abstract. The first part of this paper shows that a non-teleological account of sound change is possible if we assume two things: first, that Optimality-Theoretic constraints that do not contribute to determining the winning candidate are ranked randomly with respect to each other, i.e. differently for every speaker; second, that learners acquire as their underlying representations the forms that they detect most often in their environment. The resulting variation-and-selection scheme can be regarded as locally optimizing. It is shown, however, that it is possible that a sequence of such optimizing sound changes ends up in a loop rather than in a single absorbing final state. This kind of cyclic optimization is shown to be exactly what happened in the attested and reconstructed changes in the Indo-European consonant systems. The second part of this paper presents a simulation that shows that cyclic optimization is not only possible but also rather likely: twenty percent of all inventories are in an optimizing loop or heading towards one.

Keywords: Sound change, Optimality Theory, optimization, functional principles.

0. INTRODUCTION[1]

It is often suggested that if all sound change were due to optimizations of functional principles (minimization of articulatory effort, minimization of perceptual confusion), then sound systems should have increasingly improved during the course of history, probably to the point that they should by now have reached a stable optimum. Since the facts show, however, that sound systems tend never to stop changing, the conclusion must be, so the story goes, that optimization cannot be a major internal factor in sound change.

But it may all depend on how we define optimization. In Boersma (1989), I showed that there is a simple optimization strategy that may be cyclic, and that this cyclicity is attested in the Germanic consonant shifts. In Boersma (1997c), I showed that this optimization strategy is equivalent to a non-teleological random ranking of constraints in an Optimality-Theoretic grammar. In this chapter, I shall show that the cyclicity attested in the Germanic consonant shifts is not due to a large coincidence, but that, given random ranking of invisible constraints in OT, this cyclicity is expected in a large fraction of all sound changes.

1. ETERNAL OPTIMIZATION IS POSSIBLE

Whether a sequence of optimizations will ultimately arrive in a locally optimal state depends on how optimization is defined. Consider the following example of how not to buy a rucksack.

D. Eric Holt (ed.), Optimality Theory and Language Change, 31—65.

Table 1. Three criteria for buying a rucksack

	volume	*weight*	*price*
rucksack A	20 liters	2 kilos	€ 60
rucksack B	30 liters	4 kilos	€ 40
rucksack C	40 liters	3 kilos	€ 90

Suppose that we can choose from three rucksacks, called A, B, and C, and that we judge them on volume, weight, and price, i.e., the rucksack of our choice should be as large, light, and inexpensive as possible. Not surprisingly, the cheapest rucksack is not the largest and lightest. In fact, rucksack A is the lightest but the smallest, rucksack B is the cheapest but the heaviest, and rucksack C is the largest but the most expensive. Table 1 specifies the sizes, weights, and prices. In our decision which rucksack to buy, we will have to resolve the conflicts between the various optimization principles ("maximize volume", "minimize weight", "minimize price"). Suppose that we decide on the simplest possible decision strategy, namely that of a majority vote among the three optimization principles. Thus, we will prefer one rucksack over another if the former is better on at least two of the three points. This local decision strategy (other than a global measure of goodness) will lead to a long stay in the mountaineering shop. Suppose we are first attracted to the lightness of rucksack A, and consider buying it. We will judge, however, that there is a better alternative: when compared to A, rucksack B wins on both volume and price, so we will modify our choice in favor of B. Now that we have almost decided on buying B, we will note that rucksack C is better regarding volume and weight, so we again modify our preference, this time in favor of C. However, something now prevents us from buying C: rucksack A is better with respect to weight and price, so we again change our preference. Figure 1 shows how our decision will cycle about in a loop.

If the qualities of two rucksacks can be compared by counting the votes of optimizing principles, then each step in the loop can be regarded as an optimization of rucksack quality. The conclusion must be that if optimization is defined in this 'local' way (i.e. by comparing the current preference with an alternative candidate), then an eternally optimizing sequence of preferences is perfectly possible.

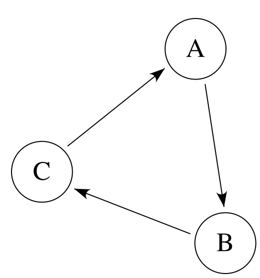

Figure 1. The simplest eternal optimization scheme.

2. OPTIMIZATION OF SOUND SYSTEMS

The teleological (i.e. goal-oriented) decision strategy of majority-vote optimization introduced above was applied by Boersma (1989) to sound change, again by using three optimizing principles ("minimize articulatory effort", "maximize perceptual contrast", "maximize perceptual salience"). Sound change, then, was modeled as follows:

Table 2. Teleological sound change

a. Start with a random phoneme inventory.
b. **Variation**: propose a randomly selected small sound change, i.e. a change of a single phoneme to a nearby phoneme.
c. **Teleological selection**: let the three functional principles vote in favor of or against this proposal.
d. Decide by a majority vote.
e. Return to step b.

Boersma (1989) showed that this model accounts for the attested cyclic behavior in the Germanic consonant shifts. The drawback of this approach, however, is the goal-orientedness in the selection step (c). Finding instead a blind underlying mechanism to account for this step would be more satisfying.

One such blind mechanism is provided by the strict-ranking decision scheme of Optimality Theory, in which it seems natural that variation can be described as a result of a set of mutually unranked constraints (Anttila 1997a). If the possible rankings within this set are distributed evenly among the population of speakers, we see the emergence of a pressure in the direction of a sound change equivalent to the results of the earlier proposal of the majority vote. Boersma (1997c) used the following variation-and-selection model for predicting the direction of sound change:

Table 3. Non-teleological sound change

a. Start with any inventory and determine its Optimality-Theoretic constraint-ranking grammar. All faithfulness constraints are ranked so high that the surface forms reflect the underlying forms perfectly.
b. **Non-variation**: with faithfulness high-ranked, the workings of many lower-ranked constraints are invisible, i.e. some constraints never contribute to determining a surface form. The mutual ranking of these invisible constraints will therefore be different for every speaker.
c. **Variation**: one faithfulness constraint, which is randomly rerankable because it refers to a non-contrastive feature, happens to fall from the top to the bottom of the entire constraint hierarchy. The formerly hidden rankings now become visible, and the speakers will reveal several new sound systems, depending on their random ranking of the originally hidden constraints.

d. **Non-teleological selection** (first reanalysis): from the pool of variation, the next generation chooses as its underlying sound system the surface system that occurs most often in this pool. This can be seen as a postponed majority decision among the speakers of the language. This generation shows the same variable surface forms as the earlier generation, but their underlying forms may be different.
e. **Freezing**: one faithfulness constraint happens to rise from the bottom to the top of the hierarchy. This causes the new underlying forms to surface faithfully.
f. **Random feature selection** (second reanalysis): now that there is no variation, the features used for underlying forms may change in the next generation.
g. Return to step b.

In the next three sections, I will discuss an example.

3. INVENTORIES OF OBSTRUENTS

As an example, I will consider symmetric inventories of labial, coronal and dorsal obstruents with three manners, in accented initial position. The labials are chosen from the set { p, b, f, v, ph }. An example of such an inventory is { ph, b, f, th, d, θ, kh, g, x }, but since it is symmetric (the same three manners are used at all three places), I will abbreviate it as { ph, b, f }.

It is important that it be made explicit what we mean if we say that an inventory is e.g. { ph, b, f }. I will take it to mean that the language has three contrasting manners of obstruents, whose phonetic realizations in accented initial position are best described as [ph], [b], and [f]. For instance, English has the labial obstruent inventory { p^h, b̥, f, v }. What the *underlying* inventory for a specific language is will have to depend on an analysis of the variation and processes within that language.

3.1. Underlying features

I will assume that the underlying representation of the three obstruents of every inventory is a subset of the features in Table 4. For later use (§5.8), I included the three breathy voiced plosives /bɦ/, /dɦ/, and /gɦ/ (= /b̤/, /d̤/, /g̤/).

Table 4. Underlying features for 15 obstruents

	p	b	f	v	ph	bɦ	t	d	θ	ð	th	dɦ	k	g	x	ɣ	kh	gɦ
voice	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+
noise	–	–	+	+	+	+	–	–	+	+	+	+	–	–	+	+	+	+
plosive	+	+	–	–	+	+	+	+	–	–	+	+	+	+	–	–	+	+
place	lab	lab	lab	lab	lab	lab	cor	cor	cor	cor	cor	cor	dor	dor	dor	dor	dor	dor

The three 'manner' features /voice/, /noise/ and /plosive/ are auditorily based, since in the view of phonology expressed here (Boersma 1998), underlying representations have been copied during the acquisition process from perceived discrete representations that the learner constructed from her auditory input. There are some differences between these perceptual features and the more traditional features [voice], [continuant] and [spread glottis] of generative phonology, which are 'hybrid' in the sense that they are at least partly based on articulations (adducted vocal folds with lax supralaryngeal musculature; incomplete closure of the oral cavity; and abducted vocal folds, respectively). The perceptual feature /voice/ corresponds to audible periodicity; for obstruents, the use of this feature is not distinguishable from the use of the hybrid voicing feature. The perceptual feature /plosive/ refers to an interruption of oral and nasal airflow. For obstruents, the values /+plosive/ and /–plosive/ correspond to the values [–continuant] and [+continuant] of generative phonology, respectively; the difference between the two features would only become apparent in the nasal stops, which are /–plosive/ but [–continuant], but we do not consider the nasals here.[2] For our purposes, the only relevant difference between the feature systems lies in the perceptual feature /noise/, which refers to audible non-periodicity. The value /+noise/ is shared by /f/, /v/, and /ph/, whereas the hybrid feature [spread glottis] would set off the aspirated /ph/ against the remaining four obstruents. Only the perceptual feature approach can thus capture the widely attested connection between aspiration and frication in sound change (e.g. alternations between [ph] and [f] or between [x/f/s] and [h]). In all, this approach firmly integrates /ph/ into the five-labial-obstruent system.

3.2. Underlying representations

I assume that the lexicon contains economical representations. For symmetric obstruent inventories with three places, one of the manner features in Table 4 will be superfluous. With respect to 'manner', therefore, the inventory { ph, b, v } can be represented underlyingly in three ways:

Table 5. Different underlyingly, but identical on the surface

	ph	b	v
/voi/	–	+	+
/noi/	+	–	+

	ph	b	v
/voi/	–	+	+
/plos/	+	+	–

	ph	b	v
/noi/	+	–	+
/plos/	+	+	–

In total, there are ten possible inventories of three labial obstruents taken from the set in Table 4. Most of these can be represented in two ways. Table 6 lists the underlying feature sets for each inventory. Generally, there is a one-to-many relation between inventories and underlying structures. The reverse is also true. The first underlying { ph, b, v } inventory in Table 5 could also represent { f, b, v }, the second could also represent { p, b, v }, and the third could equally well represent

Table 6. Possible underlying features and frequency counts for the 10 inventories

Inventory	Underlying features	Frequency (lab / cor / dor)
p b f	voi noi, voi plos	17 / 1 / 13
ph p b	voi noi	8 / 14 / 7
p b v	voi noi, voi plos	7 / 3 / 3
p f v	voi noi, voi plos	2 / 1 / 7
ph p f	noi plos	5 / 0 / 4
ph b f	voi plos, noi plos	4 / 0 / 3
ph p v	voi noi, noi plos	0 / 0 / 2
ph b v	voi noi, voi plos, noi plos	1 / 0 / 0
ph f v	voi plos	0 / 0 / 1
b f v	voi noi, voi plos	1 / 0 / 0

{ ph, p, v }, { ph, b, f }, or { ph, p, f }. This bidirectional indeterminacy will turn out to be crucial in our account of sound change.

3.3. Typology

The frequency of occurrence of our ten inventories in Maddieson's (1984) database of 317 languages is listed in the last column of Table 6, for each of the three places. For instance, { p, b, f } occurs 17 times, { t, d, θ } only once, and { k, g, x } 13 times. Bilabial fricatives were included in the counts for the labials, but sibilant fricatives were not included in the counts for the coronals.

In Table 6 we see that if a language has only one manner of plosives, this will nearly always be the plain plosive (ten times), and only rarely the aspirated plosive (once) or the voiced plosive (once). We can explain this if plosiveness tends to be the primary underlying feature, so that languages with a single plosive will prefer the one with the least effort, i.e. the one that does not involve the glottal spreading of [ph] or the supralaryngeal laxing of [b]. On the other hand, languages with a plosive inventory of { ph, b }, i.e. without the plain plosive, seem to be well attested, probably because the large auditory contrastivity of /ph/ and /b/ contributes to having much less perceptual confusion than in the case of the auditorily less contrastive pairs /ph/–/p/ or /p/–/b/. If these observations continue to hold for larger samples of languages (unfortunately, the current numbers are too small for reliable statistical tests), we will have evidence that a functionalist account in terms of minimization of articulatory effort and perceptual confusion makes better empirical predictions than generative accounts (e.g. Lombardi 1991) that attribute the preference for /p/ simply to an unexplained relative *markedness* of /ph/ and /b/, thereby ignoring the role of contrastiveness within the inventory (for a similar argument regarding a velarization–palatalization contrast, see Padgett, this volume). In the model defended in the current chapter, the rarity of certain inventories will be derived as a result, not postulated as a phonological primitive, i.e., markedness will be the *explanandum*, not the *explanans*.[3]

3.4. Variation due to a lack of contrastivity

As a criterion for free variation, we could say that segments are allowed to vary freely as long as the listener can easily reconstruct the underlying form. I will call this the *recoverability criterion*. A language with a non-variable { p, b, f } inventory can be described as a full specification of the voicing and noise features, as shown in Table 7.

Table 7. Full specification for an underlying voice-noise structure

	p	b	f
/voice/	–	+	–
/noise/	–	–	+

We note that /–voi,+noi/ not only describes /f/, but /ph/ as well. If this language is to have a non-variable { p, b, f } inventory, then either the underlying |f| has to be (weakly) specified for /–plosive/, or its /–plosive/ value will have to be inserted by the phonology (e.g. by the constraint ranking). In either case, we can say that if |f| is sometimes realized as [ph], the underlying segment can still be reconstructed by the listener. Thus, a [f]~[ph] variation is allowed if the inventory is { p, b, f } and plosiveness is the tertiary feature.

But the specification in Table 7 still seems a bit rigid. Surely a language can change at least one of the six feature values without destroying comprehension. In Table 7, either of the two non-contrastive feature values can be deleted, i.e., the voicing of |f| or the noisiness of |b| can be left unspecified, as in Table 8, where the deleted feature values have been put between parentheses.

Table 8. Allowed underspecifications for an underlying voice-noise structure

	p	b	f
/voice/	–	+	(–)
/noise/	–	–	+

and

	p	b	f
/voice/	–	+	–
/noise/	–	(–)	+

At the left-hand side of Table 8, /noise/ is the primary feature: it divides the inventory into the two parts { f } and { p, b }. The two elements of the part { p, b } are subsequently distinguished by the secondary feature /voice/.[4] An underlying |f| is now allowed to be realized as [v]. Such a language, with fully contrasting plosives but a [f]~[v] alternation, resembles Dutch, a language in which an underlying |v|, weakly specified as |+voice|, surfaces as voiceless after any obstruent. In the situation of the right-hand table, the primary feature /voice/ divides the inventory into { b } and { p, f }, whose elements are distinguished by the secondary feature /noise/. In such a language, an underlying |b| can be realized as [v]. This resembles Spanish, which has a fully contrastive set of voiceless segments { p, f, t, θ, k, x } next to the alternating voiced segments { b~β̞, d~ð̞, g~ɣ̞ }.[5] Note, however, that allowing both the [f]~[v] and the [b]~[v] alternation at the same time would cause the coalescence of underlying |f| and |b|, which would violate the criterion of recoverability.[6]

Beside the specification in Table 7, there is a second way to describe a non-variable { p, b, f } inventory, namely in terms of voicing and plosiveness features, as in Table 9.

Table 9. Full specification for an underlying voice-plosive structure

	p	b	f
/voice/	–	+	–
/plosive/	+	+	–

The /noise/ feature is now tertiary, and this allows the variation [p]~[ph], since both [p] and [ph] are realizations of a voiceless plosive. Underspecification of a secondary feature leads to the two sources of variation shown in Table 10.

Table 10. Allowed underspecifications for an underlying voice-plosive structure

	p	b	f
/voice/	–	+	(–)
/plosive/	+	+	–

and

	p	b	f
/voice/	–	+	–
/plosive/	+	(+)	–

At the left, /plosive/ is primary, and this allows a [f]~[v] alternation (again). At the right, /voice/ is the primary feature, and this again allows a [b]~[v] alternation.

We have now seen that the inventory { p, b, f } can alternate with { p, b, ph }, { p, b, v }, { p, v, f }, and { ph, b, f }, if we only count single-phoneme variations. But some of the alternations can be combined: the right-hand side of Table 8 allows { p, v, ph }, and Table 10 allows { ph, b, v } (left) and { ph, v, f } (right).

If we assume that sound changes proceed along the lines of allowed variation, we get Figure 2, in which all allowed sound changes are depicted by arrows between inventories. For simplicity, I assume that sound changes within the set of labial obstruents proceed by one phoneme at a time, so that there is no direct arrow from e.g. { p, b, f } to { ph, p, v }, { ph, b, v }, or { ph, f, v }, despite the previous paragraph. Also, direct alternations between [p] and [f] (which would otherwise be allowed to occur between /voice/-primary { p, b, v } and { f, b, v } inventories) are not considered, because [ph] lies between these two sounds.

All 15 arrows in Figure 2 are bidirectional, so that this figure does not tell us anything about the preferred directions of sound change. For that, we will have to consider in detail the Optimality-Theoretic account, whose elements will follow in the next section.

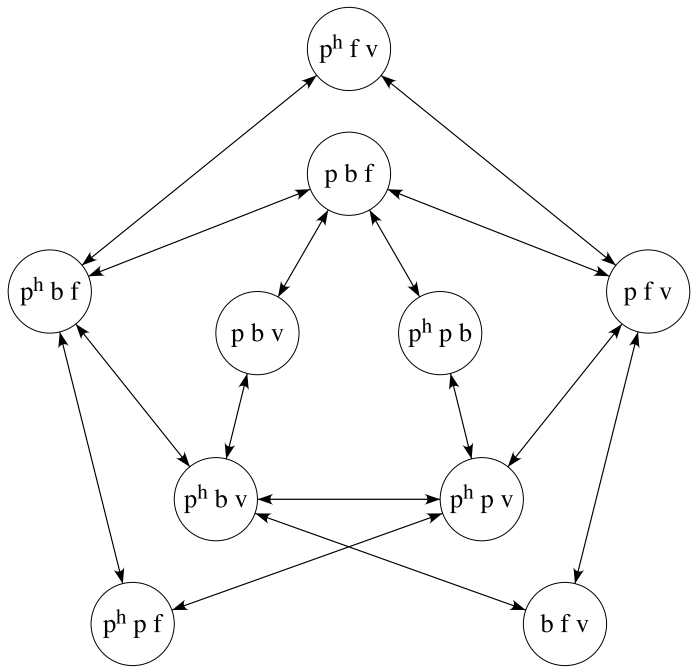

Figure 2. The fifteen considered sound changes among the ten possible inventories.

4. FIXED RANKINGS IN OBSTRUENT SYSTEMS

In this section, I will translate feature specifications into Optimality-Theoretic constraint rankings, and discuss which rankings can be regarded as fixed, and which must be language-dependent.

According to the theory of Functional Phonology (Boersma 1998), a large number of articulatory constraints, which evaluate continuous phonetic implementations, sit happily together in the production grammar with a large number of faithfulness constraints, which evaluate the similarity of discrete perceptual results to the underlying forms. In this theory, no other production constraints appear necessary to describe the typological facts. Specifically, the traditional Optimality-Theoretic markedness constraints introduced by Prince & Smolensky (1993) turn out to be superfluous, since all effects traditionally associated with markedness derive from the interaction of articulatory and faithfulness constraints alone. For the case at hand, this will be shown in §5.7.

According to Prince & Smolensky's (1993) concept of *harmonic ordering*, some Optimality-Theoretic constraint families can be internally ranked in a language-independent way. Analogously, Functional Phonology proposes a set of *local-*

ranking principles, according to which the members of constraint families can be ranked on the basis of the relative extent to which they satisfy functional principles. For our set of obstruents, the fixed rankings are listed in Table 11.

Table 11. Functional principles that lead to fixed rankings for obstruents

a. Minimization of articulatory effort yields a single fixed hierarchy of articulatory constraints (§4.1).
b. Maximization of the perceptual place contrast yields five fixed hierarchies of perceptual place faithfulness constraints (§4.2).
c. Maximization of the perceptual voicing contrast yields five fixed hierarchies of perceptual voice faithfulness constraints (§4.3).
d. Maximization of the perceptual noisiness contrast yields five fixed hierarchies of perceptual noise faithfulness constraints (§4.4).
e. Maximization of the perceptual plosiveness contrast yields five fixed hierarchies of perceptual plosive faithfulness constraints (§4.5).

The hierarchies in the following five sections are adapted from Boersma (1997c).

4.1. Fixed hierarchy for articulatory effort

The production grammar contains a single large family of articulatory constraints:

*GESTURE (*articulator*: *gesture* / *distance*, *duration*, *velocity*, *precision*):
A certain *articulator* (or combination of articulators) does not perform a certain *gesture*, over a certain *distance*, during a certain *duration*, and with a certain *velocity* and *precision*.

According to the local-ranking principle for articulatory constraints (Boersma 1998:160), articulatory constraints for the same gesture can be ranked in a fixed way on the basis of articulatory effort, if they differ in a single argument. Thus, *GESTURE is ranked higher if the distance, duration, velocity or precision is greater and everything else stays equal.

Consider, for instance, the glottal spreading gesture (posterior cricoarytenoid activity) associated with devoicing. The articulatory form [pha] must be more difficult in this respect than [pa] or [fa], since if voicelessness is called for, the active glottal spreading gesture must be stronger if the supralaryngeal vocal tract is unimpeded, as in the aspiration phase of [pha], than if the oral and nasal cavities are wholly or nearly sealed off, as during the closure periods of [pa] and [fa]. We can express this as the continuous constraint family *GESTURE (glottis: spread / distance ≥ x), which has a fixed partial ranking of *GESTURE (glottis: spread / distance ≥ 3 mm) » *GESTURE (glottis: spread / distance ≥ 2 mm) » *GESTURE (glottis: spread / distance ≥ 1 mm). For readability, I will write these constraints as "glot < [ph]" » "glot < [f]" » "glot < [p]", where "glot < [x]" is an abbreviation for "do not perform a glottal spreading gesture at least as difficult as that required for a typical [*x*]". Similarly, breathy voicing, as in [bɦ], requires less glottal spreading than [ph], which leads to "glot < [ph]" » "glot < [bɦ]". Note that these notations do not imply that the speaker has any knowledge about the discrete symbols [ph], [f], and [bɦ]; she only knows about continuous degrees of glottal spreading, and the three symbols are only there for the sake of the linguist who is reading these lines.

We can posit a similar hierarchy of anti-precision constraints. I refer here to the precision required for producing a constriction suitable for frication. If /v/ is allowed to be sometimes pronounced as the approximant [ʋ], and /f/ always has to be pronounced as a fricative, the required precision will be greater for the typical [f] than for the typical [v], so we have the fixed ranking "prec < [f]" » "prec < [v]" » "prec < [p,b,ph]".[7]

Finally, we can posit a hierarchy of constraints against the gesture needed to make an obstruent voiced, perhaps by laxing the walls of the supralaryngeal vocal tract. Since voicing requires the maintenance of glottal airflow, the effort will be higher for stronger constrictions, leading to the fixed hierarchy "lax < [b]" » "lax < [v]" » "lax < [p,ph,f]".

In Boersma (1989, 1997c), these fixed rankings were simplified to the hierarchy in Figure 3, to which I have now added the voiced aspirate [bɦ].

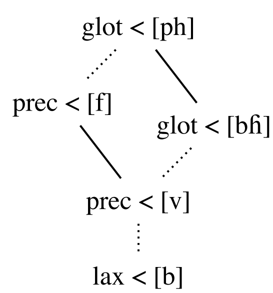

Figure 3. Hierarchy of articulatory constraints.

In this picture, the two universal rankings are given by solid lines. According to the local-ranking principle, the three other rankings must be language-dependent, and that is why I represent them by dotted lines. For the purposes of this chapter, however, I keep them fixed in order to suggest the idea that sound change is inspired by a global rather than a local measure of effort. This reflects the idea that global effort measures can predict that in the pool of variation, constraints against more effortful gestures tend to be high ranked more often than constraints against less effortful gestures.

According to the OT maxim of *factorial typology*, there must be languages in which all 'manner' faithfulness constraints are ranked low, so that the articulatory constraints have all the say in the matter. This will yield a language with a single labial obstruent, which according to Figure 3 must be [p], as we see in Tableau 1.

Tableau 1. The pronunciation of the only labial obstruent

|labial, +voice, +noise, –plosive|	glot < [ph]	prec < [f]	glot < [bɦ]	prec < [v]	lax < [b]	FAITH (voice)	FAITH (noise)	FAITH (plosive)
[ph]	*!		*			*		*
☞ [p]						*	*	*
[b]					*!		*	*
[v]				*!				
[f]		*!		*		*		
[bɦ]			*!		*			*

Each of the five losing articulatory candidates violates its 'own' gestural constraint; two of these candidates also violate a lower-ranked gestural constraint a fortiori, and [bɦ] violates two gestural constraints because it involves both a spreading and a laxing gesture. Even while the underlying form in Tableau 1 has all features specified for a typical /v/, which should be possible according to the OT maxim of *richness of the base*, the result is a plain [p], violating three faithfulness constraints. Of course, the OT device of *lexicon optimization*, which minimizes faithfulness violations in comprehension, will cause the underlying form to become |labial, –voice, –noise, +plosive|, or, which is more likely, a simple |labial|, underspecified along the lines of §3.2 because none of the three features is contrastive. Within the theory of Functional Phonology, the three faithfulness constraints would not even exist, because the acquisition device does not create faithfulness constraints that refer to features that are not perceived, and the perception grammar will not perceive features that do not underlyingly occur in the lexicon. A more interesting case, with underlying contrasts and specifications, will involve *active* faithfulness constraints, and these will be discussed in the next four sections.

4.2. Fixed hierarchy for faithfulness of perceptual place

Beside the gestural constraints, which evaluate continuous articulations, the production grammar contains several families of faithfulness constraints that evaluate the perceived similarity of a surface form to the underlying form. The only family that concerns us here is *REPLACE, which compares the perceived phonological feature values with those in the underlying representation:

> *REPLACE (*feature*: *x*, *y* / *condition* / *probability*):
> Do not realize a value *x* of an underlying perceptual *feature* as something that the listener will perceive (under a certain *condition* and with a certain *probability*) as a different value *y* of that same feature.

According to the local-ranking principle for faithfulness constraints (Boersma 1998:177), these constraints are ranked higher if their violation would cause more confusion. I will consider three universal local hierarchies.

The first hierarchy to be considered is that for perceptual place. Labiality faithfulness constraints indirectly express the desire to keep the labial obstruents perceptually distinctive from the coronal and velar obstruents. One of the many labiality faithfulness constraints is *REPLACE (place: lab, cor / *prob* > 20%), which states that an underlying labial segment should not be pronounced as something that has a probability of over 20% of being perceived as a coronal. If this 20% happens to be the probability that a typical [ba] is perceived as /da/, then we can write this *REPLACE constraint more legibly as "lab (|lab|) ≥ [ba]", which is an abbreviation of "for a segment specified underlyingly as |labial|, the auditory cues for labiality should be at least as good as the cues available in a typical [ba]". The labiality cues associated with a typical [va] tend to be worse than those associated with a typical [ba], if we take into account the ubiquity with which fricatives change place through history. So, having [v]-like place cues causes more perceptual confusion than having [b]-like place cues, and this is expressed as a fixed ranking like *REPLACE (place: lab, cor / *prob* > 30%) » *REPLACE (place: lab, cor / *prob* > 20%). In legible notation, the constraint "lab (|lab|) ≥ [va]" must outrank "lab (|lab|) ≥ [ba]". It is also likely that voicing obscures the place cues, so we must assume that the pair /b/–/g/ is more confusable than /p/–/t/, and /v/–/ɣ/ is more confusable than /f/–/x/. Finally, the place cues may be best in the least voiced environment, i.e. for the aspirated pair /ph/–/kh/. This leads to the labiality hierarchy in Figure 4.

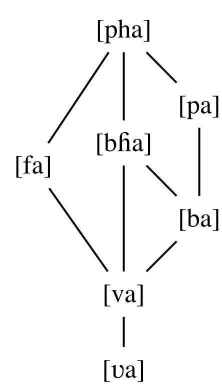

Figure 4. Labiality hierarchy.

The grammar in Figure 5 translates the labiality hierarchy into partial constraint hierarchies for our five labial segments. In this figure, the generic "lab (|lab|)" constraint has been divided into constraints for the five separate segments. Thus, the constraint "lab (|b|) ≥ [b]" states that an underlying |b| should be at least as labial as a typical [b].

lab (|p|) ≥ [b]
|
lab (|p|) ≥ [p]
|
lab (|p|) ≥ [ph]

lab (|ph|) ≥ [p]
|
lab (|ph|) ≥ [ph]

lab (|b|) ≥ [v]
|
lab (|b|) ≥ [b]
|
lab (|b|) ≥ [p]

lab (|f|) ≥ [v]
|
lab (|f|) ≥ [f]
|
lab (|f|) ≥ [ph]

lab (|v|) ≥ [ʋ]
|
lab (|v|) ≥ [v]
|
lab (|v|) ≥ [f]

Figure 5. Fixed hierarchies for place faithfulness.

Again, the shorthand notation does not imply that the speaker has any knowledge about the symbols [b], [v], [ʋ], [f], [p], or [ph]: she only knows about confusability, and the symbols are only here for the sake of the reader of these lines.

4.3. Fixed hierarchies for voice faithfulness

Analogous hierarchies can be posited for manner features. A segment specified underlyingly for /+voice/ should not be pronounced as something that runs a high risk of being perceived as /–voice/, at least if it has to contrast with a voiceless segment. As in the previous section, we can translate this as: a segment specified for /+voice/ should be pronounced with as many voicing cues as possible. The underlying segment |b|, for instance, which is shorthand for /voiced labial plosive/, should preferably surface as the most voiced plosive, i.e. the implosive [ɓ], or, if that is not possible, it should have the voicing of a typical prevoiced [b], and if that is not possible either, it should certainly be as voiced as the lenis voiceless [b̥]. This leads to the universal hierarchy “voi (|b|) ≥ [b̥]” » “voi (|b|) ≥ [b]” » “voi (|b|) ≥ [ɓ]”. Again, this notation stands for the less legible *Replace (voice: +, – / plosive / *prob*) family, e.g., a [ɓ] pronunciation is less likely to be mistaken for something voiceless than a [b] pronunciation is; the speaker’s knowledge only concerns confusability, not the discrete symbols [b̥], [b], or [ɓ]. An analogous hierarchy can be posited for the voiced labial fricative and for the three voiceless segments. Figure 6 shows the degree of voicing for nine labial obstruents.

Figure 7 shows the five universal hierarchies that can be derived from the voicedness hierarchy for our set of five obstruents. The solid lines depict the fixed rankings, and the five hierarchies are freely ranked with respect to each other, e.g., “voi (|f|) ≤ [ph]” could outrank “voi (|b|) ≥ [p]” in some languages. I have restricted the symbols for the degrees of voicing to the set that we will consider in this chapter; this ignores the implosives, and regards the lenis voiceless plosives as plain

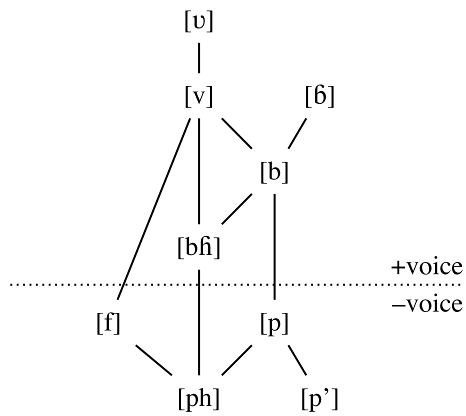

Figure 6. Voicedness hierarchy.

voiceless. The interpretation of the rankings in Figure 7 with respect to variation and change is as follows. Usually, “voi (|b|) ≥ [b]” will be ranked high enough to ensure that an underlying |b| will surface as [b]. If not, the constraint “voi (|b|) ≥ [p]” may still be ranked highed enough to ensure that the underlying |b| will only minimally change, i.e. to [p], and not all the way to e.g. [ph]. The constraint “voi (|b|) ≥ [v]” expresses the preference for |b| to be pronounced even more voiced, like a typical [v]; if this constraint is allowed to override some other constraints, |b| is realized with voicing enhancement, i.e. as [v]. In this chapter, the ‘extra high’ constraints such as “voi (|b|) ≥ [p]” will be considered unviolable, so that we will need to consider changes to adjacent segments only (i.e., |b| can go to /p/ and /v/, but not to /ph/ or /f/). The rankings of ‘extra low’ constraints such as “voi (|b|) ≥ [v]”, by contrast, will be seen to play large roles in determining variation and change.

voi (|p|) ≤ [b]	voi (|ph|) ≤ [p]	voi (|b|) ≥ [p]
voi (|p|) ≤ [p]	voi (|ph|) ≤ [ph]	voi (|b|) ≥ [b]
voi (|p|) ≤ [ph]		voi (|b|) ≥ [v]

voi (|f|) ≤ [v]	voi (|v|) ≥ [f]
voi (|f|) ≤ [f]	voi (|v|) ≥ [v]
voi (|f|) ≤ [ph]	voi (|v|) ≥ [ʋ]

Figure 7. Fixed hierarchies for voice faithfulness.

4.4. Fixed hierarchies for noise faithfulness

Segments specified underlyingly for /+noise/, i.e. fricatives and aspirated plosives, should be pronounced with the best noisiness cues possible. The voiceless fricative [f] will be the noisiest segment, because for the other segments the noise cues are reduced either by voicing, as in the fricative [v] and the aspirated plosive [bɦ], or by a near-silent period, as in [ph] and [bɦ]. Segments specified underlyingly for /–noise/, i.e. glides and non-aspirated plosives, should be pronounced with the best non-noisiness cues possible. The plosives will be less noisy than the glide [ʋ]. All this leads to the hierarchy in Figure 8.

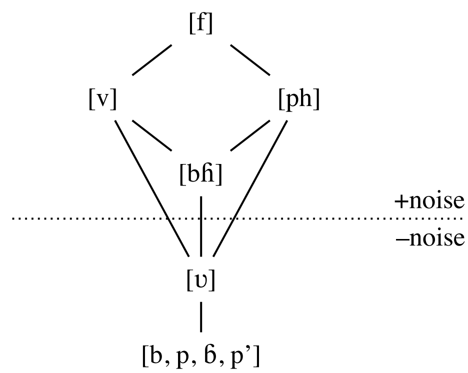

Figure 8. Noisiness hierarchy.

The five constraint hierarchies in Figure 9 are constructed from this.

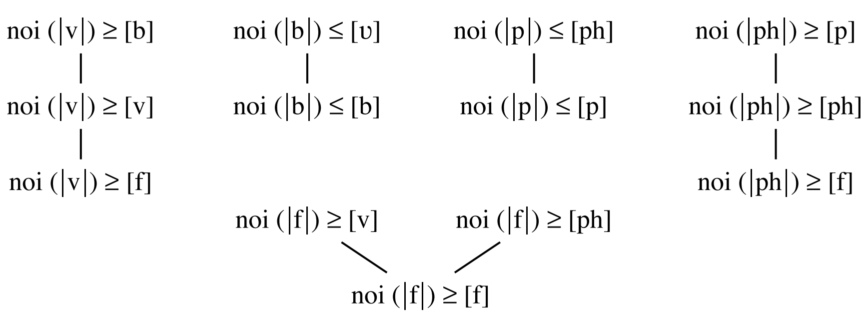

Figure 9. Fixed hierarchies for noise faithfulness.

4.5. Fixed hierarchies for plosive faithfulness

Finally, the five segments divide into three plosives (ph, p, b) and two fricatives (f, v). If we assume that fricatives are worse plosives than plosives are, and that voicing and aspiration weaken the plosiveness of a plosive while ejection strengthens it, we get the plosiveness or continuancy hierarchy in Figure 10.

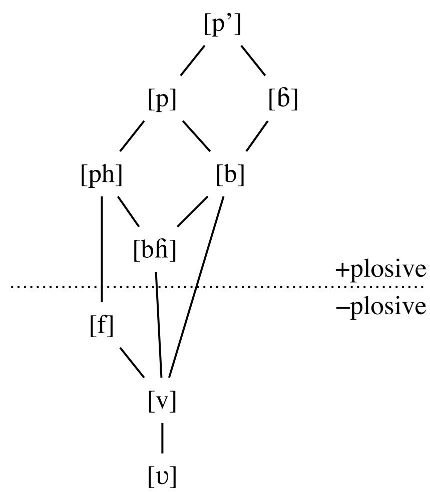

Figure 10. Plosiveness hierarchy.

The corresponding fixed constraint rankings are shown in Figure 11.

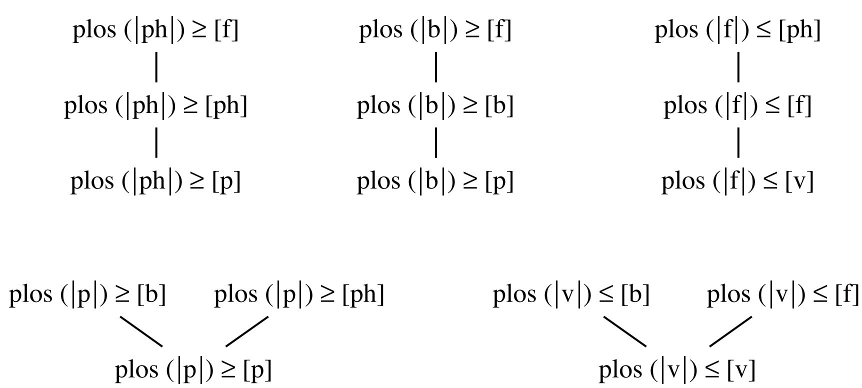

Figure 11. Fixed hierarchies for plosive faithfulness.

5. A CIRCULAR SOUND CHANGE

This section will describe in detail how half of the { p, b, v } inventories tend to change towards { p, b, f } under the variation-and-selection model of §2 and given the fixed rankings of §4. I will generalize this example to another 14 possible changes within the set of three-obstruent inventories (or even 16, if a couple of ‘double’ changes are included), showing that the complete set of changes amounts to a circular optimization similar to the rucksack example of §1.

5.1. First generation: a non-varying { p, b, v } language

An Optimality-Theoretic constraint grammar for a { p, b, v } language based on the features /voice/ and /noise/ is shown in Figure 12.

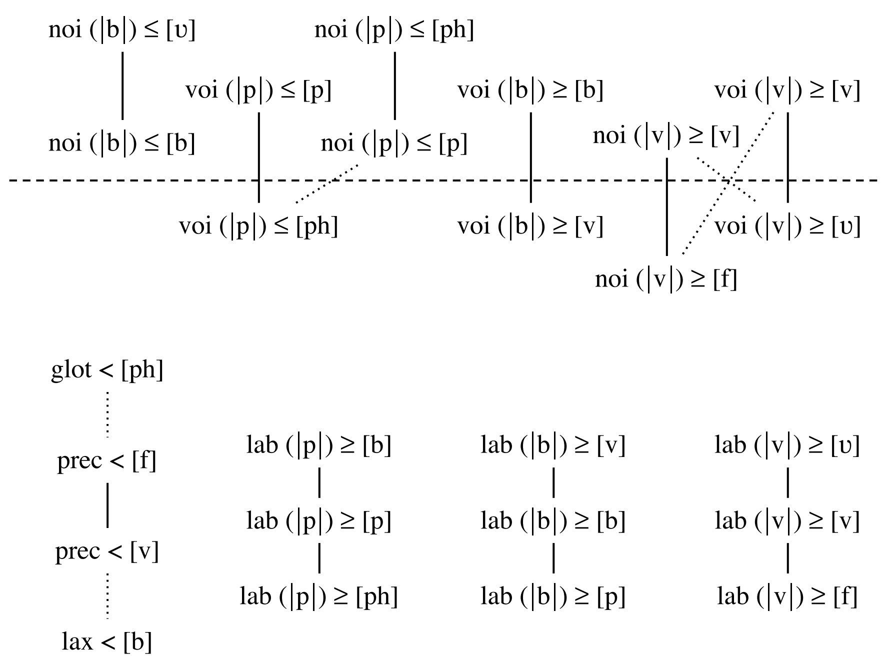

Figure 12. Grammar of a non-varying { p, b, v } language.

In this grammar, we see only ten of the 21 fixed hierarchies of §4. The 11 remaining hierarchies do not appear in the grammar of this language, since they refer to representations that cannot occur in the lexicon. This is a maxim of Functional Phonology: the lexicon contains only those representations that the listener has learned to perceive, and the listener perceives only those features that are needed for economical lexical representation (this explicitly goes against Prince & Smolensky’s notion of Richness of the Base). Hence, faithfulness constraints, which compare lexical representations with perceived forms, will never refer to non-lexical features. Thus, the faithfulness hierarchies for /plosive/ do not appear in the grammar, since this feature is superfluous. Also, the faithfulness hierarchies for |f| and |ph| do not appear in the grammar, since these are not underlying segments.[8]

For reasons of exposition, I have divided the constraints in Figure 12 into two *strata*. The constraints above the dashed line can be said to be in the first (highest) stratum, since none of them is ever violated in surface forms. The constraints below the line are in the second stratum; they are dominated by the constraints above the line, but are not ranked with respect to each other (except for the fixed rankings). The following tableaux show how the three underlying segments are realized. Tableau 2 shows that an underlying |p|, i.e. "voiceless non-noisy" according to Table 7, is pronounced as [p]. A double vertical line marks the virtual boundary between the two strata.

Tableau 2. The pronunciation of the voiceless non-noisy obstruent

|p| i.e. /–voice, –noise/	noi (|p|) ≤ [ph]	noi (|p|) ≤ [p]	voi (|p|) ≤ [p]	voi (|p|) ≤ [ph]	lab (|p|) ≥ [ph]
[ph]		*!			
☞ [p]				*	*
[b]			*!	*	*
[v]	*!	*	*	*	*
[f]	*!	*	?	*	*
[ʋ]		*!	*	*	*

The two fricative candidates are ruled out because they are too noisy: they violate the constraint that says that an underlying |p| should certainly not produce more noise than a typical [ph] does (remember that this constraint represents knowledge about noisiness, not about [ph]). The candidate [ph], which may occur in the candidate list if the speaker has learned a sensorimotor mapping for it, is a bit too noisy as well. The candidates [b], [v], and [ʋ] are too voiced: they violate a constraint against voicing an underlying |p|. Note that "noi (|p|) ≤ [p]", i.e. a constraint above the line in Figure 12, crucially outranks "voi (|p|) ≤ [ph]", a constraint below the line. Otherwise, [ph] would have been the winner. Several crucial rankings like this one show up as dotted lines in Figure 12.

In Tableau 3, we see that an underlying |b|, i.e. "voiced non-noisy", is pronounced as [b].

Tableau 3. The pronunciation of the voiced non-noisy obstruent

|b| i.e. /+voice, –noise/	noi (|b|) ≤ [ʋ]	noi (|b|) ≤ [b]	voi (|b|) ≥ [b]	voi (|b|) ≥ [v]	lax < [b]
[ph]	*!	*	*	*	
[p]			*!	*	
☞ [b]				*	*
[v]	*!	*			
[f]	*!	*	*	*	
[ʋ]		*!			

In Tableau 4, we see that an underlying |v|, i.e. "voiced noisy", is pronounced as [v].

Tableau 4. The pronunciation of the voiced noisy obstruent

|v| i.e. /+voice, +noise/	noi (|v|) ≥ [v]	voi (|v|) ≥ [v]	voi (|v|) ≥ [ʋ]	noi (|v|) ≥ [f]	prec < [v]
[ph]	?	*!	*	*	
[p]	*!	*	*	*	
[b]	*!	*	*	*	
☞ [v]			*	*	*
[f]		*!	*		*
[ʋ]	*!			*	

No other candidate than [v] is noisy and voiced enough to beat it: [ph] may be as noisy as a typical [v] (see Fig. 8), but it is not voiced at all; the noisiness of [f] is more than enough, but it is not voiced either; [b] is voiced, but less than a typical [v]; the voicing of [ʋ] is more than enough, but it is not noisy. Note that it is crucial that first-stratum "noi (|v|) ≥ [v]" outranks second-stratum "voi (|v|) ≥ [ʋ]"; otherwise, candidate [ʋ] would win. It is also crucial that first-stratum "voi (|v|) ≥ [v]" outranks second-stratum "noi (|v|) ≥ [f]"; otherwise, [f] would win.

5.2. Second generation: a varying { p, b, v } language

The grammar in Figure 12 is rather rigid. Considerations of perceptual recoverability (§3.4) lead us to identify the two types of variation shown in Table 12.

Table 12. Allowed underspecifications for an underlying voice-noise structure

	p	b	v
/voice/	–	+	(+)
/noise/	–	–	+

and

	p	b	v
/voice/	–	+	+
/noise/	(–)	–	+

As an example, we consider a simplified form of Dutch, in which [f] is a positional variant of |v|, which is devoiced after any obstruent. The allowed underspecification, therefore, is as on the left side of Table 12. This underspecification, however, is a bit too strong: an underlying |v| is not totally unspecified for voicing. Instead, |v| wants to surface as voiced, but it will give up this desire if stronger forces require it to be pronounced as [f]. Therefore, it is more appropriate to regard |v| as *weakly specified* for /+voice/. An Optimality-Theoretic account in terms of our fixed rankings shows exactly this property if the constraint "voi (|v|) ≥ [v]" is ranked low, so that an articulatory constraint against voiced fricative-final obstruent clusters can overrule the /+voice/ specification of |v| and force it to surface as [f]. Tableau 5 shows that even with a low-ranked [+voice] specification, |v| will normally end up as voiced, as long as "voi (|v|) ≥ [v]" outranks some constraints for maximization of labiality and noisiness.

Tableau 5. Postvocalic pronunciation of the voiced noisy obstruent

|ava| i.e. /+voice, +noise/	*[voiced fricative / obstruent _]	voi (|v|) ≥ [v]	lab (|v|) ≥ [f]	noi (|v|) ≥ [f]
☞ [ava]			*	*
[afa]		*!		

The constraint "*[voiced fricative / obstruent _]" stands in for the *GESTURE constraint that disallows post-obstruent voiced fricatives; it is a typical OT constraint, in that it could be satisfied by devoicing the fricative, deleting or gliding the obstruent, gliding the fricative, or inserting a vowel. In Dutch, the most drastic of these options will be prevented by high-ranked faithfulness constraints, so that post-obstruent |v| will be devoiced, as shown in Tableau 6.

Tableau 6. Post-obstruent pronunciation of the voiced noisy obstruent

|atva| i.e. /+voice, +noise/	*[voiced fricative / obstruent _]	voi (|v|) ≥ [v]	lab (|v|) ≥ [f]	noi (|v|) ≥ [f]
[atva]	*!		*	*
☞ [atfa]		*		

This exercise about positional neutralization is included in order to show that the same kind of constraints account for the ‘phonetic’ choices of articulation as well as for ‘phonological’ processes like assimilation.

In the example of Tableaux 5 and 6, “voi (|v|) ≥ [v]” was ranked at an intermediate height. But according to the criterion of perceptual recoverability (§3.4), the variation between [v] and [f] realizations could be completely free, i.e., “voi (|v|) ≥ [v]” could be ranked very low, perhaps in a third stratum. This extreme version of Dutch is shown in Figure 13.

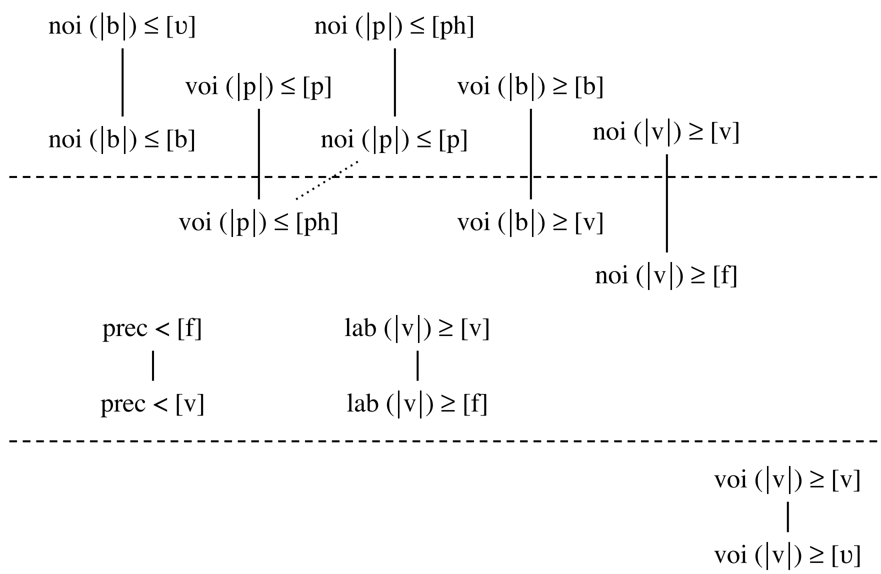

Figure 13. Varying { p, b, v } language.

Now that the /+voice/ specification for |v| has dropped to the bottom of the hierarchy, the surface form will be determined by the ranking of the constraints in the second stratum, which used to be invisible in the previous generation, as was shown in Figure 12. There are three relevant constraints here: “noi (|v|) ≥ [f]”, “prec < [f]”, and “lab (|v|) ≥ [f]”.

These three constraints are ranked in an unpredictable order in the pool of between-speaker variation. If “noi (|v|) ≥ [f]” happens to be ranked on top of these three, the noisiness contrast of |v| with respect to |b| and |p| will be enhanced by pronouncing it as [f], as shown in Tableau 7.

Tableau 7. Enhancement of noisiness contrast

|pabava|	noi (|v|) ≥ [f]	prec < [f]	lab (|v|) ≥ [f]	voi (|v|) ≥ [v]
[pabava]	*!		*	
☞ [pabafa]		*		*

If “prec < [f]” happens to be on top, the input will surface faithfully, as shown in Tableau 8.

Tableau 8. Minimizing precision

|pabava|	prec < [f]	noi (|v|) ≥ [f]	lab (|v|) ≥ [f]	voi (|v|) ≥ [v]
☞ [pabava]		*	*	
[pabafa]	*!			*

And if “lab (|v|) ≥ [f]” happens to be on top, the place contrast of |v| with respect to other fricatives such as |ð| and |ɣ| will be enhanced, as shown in Tableau 9.

Tableau 9. Enhancement of place contrast

|pabava|	lab (|v|) ≥ [f]	noi (|v|) ≥ [f]	prec < [f]	voi (|v|) ≥ [v]
[pabava]	*!	*		
☞ [pabafa]			*	*

If all three constraints have an equal probability of being ranked on top in the pool of between-speaker variation, two-thirds of the speakers will devoice an underlying |v|.

This section has shown that the maximum free variation in OT is achieved with random reranking of intermediate constraints, keeping directly or indirectly contrastive specifications fixed at the top and redundant specifications fixed at the bottom.

5.3. Third generation: reanalysis to { p, b, f }

The third generation hears [pabafa] more often than [pabava], so they construct |pabafa| as the underlying form, i.e., their fricative segment is specified as /–voice/. The result is a change from underlying |pabava| to underlying |pabafa| in two generations.

One would think that the reanalysis step does not lead to a change in the surface forms. After all, a voiceless specification constraint “voi (|f|) ≤ [f]” will now be ranked in the bottom stratum (taking the place of the now defunct “voi (|v|) ≥ [v]”), resulting in one-third [v] realizations. However, if this constraint does go up in the grammar, for whatever reason, the surface form will become [f] 100 percent of the time, and the overall result will be that the surface inventory has changed from a non-variable { p, b, v } to a non-variable { p, b, f } in three steps. This account may seem unsatisfactory, but we should note that both the fall of “voi (|v|) ≥ [v]” and the rise of “voi (|f|) ≤ [f]” can be seen as random changes in the ranking of faithfulness constraints whose ranking is immaterial to comprehension. Thus, the changes in the rankings of these constraints have no direction; they go up and down the hierarchy. The result of these random movements, though, is an irreversible directional sound change from [v] to [f]. The situation is analogous to the working of most combustion engines, which convert an up-and-down motion into a cyclic motion.

5.4. From elsewhere to { p, b, f }

The inventory { p, b, f } is a very good one: if the specified features are primary /voice/ and secondary /noise/, this language will not change any further. It cannot go back to { p, b, v }, since the same three constraints that caused the change from { p, b, v } to { p, b, f } will now vote against the reversal; in general, changes are unidirectional if the underlying representations do not change. So what happens if we reverse the primacy of /voice/ and /noise/ in a { p, b, f } inventory? The voiceless obstruents |p| and |f| will now contrast for /noise/, but the voiced obstruent |b| may be underspecified for /noise/, which means that the constraint “noi (|b|) ≤ [b]” may drop to the third stratum. Tableaux 10 show the three relevant rankings that are then left in the second stratum.

Tableaux 10. Struggle between { p, b, f } and { p, f, v }

|pabafa|	voi (|b|) ≥ [v]	prec < [v]	lab (|b|) ≥ [b]	noi (|b|) ≤ [b]
[pabafa]	*!	*		
☞ [pavafa]		**	*	*

|pabafa|	prec < [v]	voi (|b|) ≥ [v]	lab (|b|) ≥ [b]	noi (|b|) ≤ [b]
☞ [pabafa]	*	*		
[pavafa]	**!		*	*

|pabafa|	lab (|b|) ≥ [b]	voi (|b|) ≥ [v]	prec < [v]	noi (|b|) ≤ [b]
☞ [pabafa]		*	*	
[pavafa]	*!		**	*

We see that two of the three drives (maximization of labiality and minimization of precision) prefer the { p, b, f } inventory, which will therefore not change into { p, f, v }. On the contrary, Tableaux 10 imply that { p, f, v } inventories can become { p, b, f }.

There is a third inventory that can become { p, b, f }. It is { ph, p, b }, which always has underlying /noise/ and /voice/, since /plosive/ can never distinguish between any elements of this inventory. If this inventory is non-varying, there must be something that guarantees that |ph| is realized as [ph] rather than [f], which has the same representation. In a derivational account, we could achieve this either by including the tertiary feature value /+plosive/ in the feature specification table, or by having a default rule supply this value. In our constraint-based account, the phonetic realization is simply guaranteed by a high ranking of the constraint "voi (|ph|) ≤ [ph]". But because it does not contribute to contrastiveness, this constraint may drop down the hierarchy. For an underlying |phabapa|, the faithful candidate [phabapa] is preferred only by "lab (|ph|) ≥ [ph]", whereas its competitor [fabapa] is preferred by "noi (|ph|) ≥ [f]" and "glot < [ph]" (assuming that this outranks "prec < [f]"). This spirantization of aspirates is what probably happened in Proto-Latin (compare Classical Greek /pheroː/ 'I carry' with Latin /feroː/).

Thus, all of the three inventories { p, f, v }, { p, f, v }, and { ph, p, b } can change to { p, b, f }. It almost looks as though this most common inventory is a *sink* for sound change, i.e. an absorbing state that a language cannot get out of.

5.5. Out of { p, b, f }

But there are two ways to escape the favored { p, b, f } inventory. As seen in Table 10 (left), a non-alternating { p, b, f } may be reanalyzed by the learners in terms of primary /plosive/ (distinguishing |p| and |b| from |f|) and secondary /voice/ (distinguishing |p| from |b|), so that |f| is now non-contrastively specified for /–voice/. Surprisingly, the language can then return to { p, b, v }, as shown in Tableaux 11.

Tableaux 11. Struggle between { p, b, f } and { p, b, v }

|pabafa|	plosive (|f|) ≤ [v]	prec < [f]	lab (|f|) ≥ [f]	voi (|f|) ≤ [f]
☞ [pabava]			*	*
[pabafa]	*!	*		

|pabafa|	prec < [f]	plosive (|f|) ≤ [v]	lab (|f|) ≥ [f]	voi (|f|) ≤ [f]
☞ [pabava]			*	*
[pabafa]	*!	*		

|pabafa|	lab (|f|) ≥ [f]	plosive (|f|) ≤ [v]	prec < [f]	voi (|f|) ≤ [f]
[pabava]	*!			*
☞ [pabafa]		*	*	

There is only a single difference if we compare these tableaux with Tableaux 7–9. It is the difference between the features /noise/ and /plosive/: [f] is noisier than [v], but [v] is more continuant than [f]. We now see that { p, b, f } can change into { p, b, v }; this must have happened at some time in the history of Dutch.

The other way for { p, b, f } to change is to realize |p| as [ph]. This can happen if the *primary*-feature constraint "plos (|p|) ≥ [p]" drops down the hierarchy while the higher-ranked "plos (|p|) ≥ [ph]" does not. This partial dropping of faithfulness is allowed by the recoverability criterion (§3.4), but I will assume that these changes only occur in cases where a lowering of secondary-feature constraints would not predict the opposite change. Once "plos (|p|) ≥ [p]" has dropped, the constraints "voi (|p|) ≤ [ph]" and "lab (|ph|) ≥ [ph]" will prefer the [ph] variant, and only "glot < [ph]" will prefer [p]. The next generation will reanalyze the resulting interspeaker variation, in which [ph] is more common than [p], as an underlying { ph, b, f } inventory.

5.6. Predicted possible sound changes

We just handled five changes; for a detailed analysis of nine more, see Boersma (1997c). In total, our variation-and-selection scheme predicts 24 preferred changes, all of which are listed in Table 13.

The example of §5.1–3 is summarized in the seventh row of Table 13: the change is from { p, b, v } to { p, b, f }, the feature tree had /noise/ as its primary branching and /voice/ as its secondary branching, the constraints that voted in favor of the change were noise faithfulness and place faithfulness, and the constraint that voted against the change was minimization of precision. The table is divided into four smaller tables: the 12 entries at the top show changes in the sign of the secondary feature (§5.2, §5.4, §5.5); the next eight entries show changes in the realization of the secondary feature (§5.4); the next two entries show changes in the secondary contrast (as suggested in §5.5, we use them here for lack of a stronger way to determine the direction of the arrows between { ph, p, v }–{ p, f, v } and { p, b, f }–{ ph, b, f }); finally, the two changes between parentheses are changes in (the realization of) two features at the same time. The table predicts only one reversible change: that between { p, b, f } and { p, b, v }. The term "lax" in the table stands for the constraint "lax < [b]", which is satisfied by pronouncing |b| as [p].

Figure 14 shows the 22 possible 'single' system changes as 15 solid lines between ten surface inventories, one of which is bidirectional. The dotted lines involve the segment |bɦ| and will be discussed in §5.8.

Table 13. Preferred changes in inventories of three labial obstruents excluding |bɦ|

From:	To:	Features:	In favor:	Against:
ph p v	ph p b	voice (noise)	lab, prec-v	+voi
ph b f	ph b v	plosive (voice)	–plos, prec-f	lab
ph b v	ph p v	noise (voice)	lab, lax	
ph f v	ph f b	voice (plosive)	lab, prec-v	+voi
p b f	p b v	plosive (voice)	–plos, prec-f	lab
p b v	ph b v	voice (noise)	–voi, lab	glot
p b v	p b f	noise (voice)	+noi, lab	prec-f
p f v	p f b	voice (noise)	lab, prec-v	+voi
p f v	p f b	voice (plosive)	lab, prec-v	+voi
b f v	b ph v	voice (plosive)	–voi, lab	glot
b f v	p f v	noise (voice)	lab, lax	
b f v	p f v	plosive (voice)	+plos, lab, lax	
ph p b	f p b	noise (voice)	+noi, glot	lab
ph p f	ph p v	plosive (noise)	–plos, prec-f	lab
ph b f	ph p f	noise (plosive)	lab, lax	
ph b f	ph b v	plosive (noise)	–plos, prec-f	lab
ph b v	ph p v	noise (plosive)	lab, lax	
ph f v	p f v	plosive (voice)	+plos, glot	lab
p b v	ph b v	voice (plosive)	–voi, lab	glot
b f v	b ph v	voice (noise)	–voi, lab	glot
ph p v	f p v	voice (noise)	+noi, glot	lab
p b f	ph b f	plosive (voice)	–voi, lab	glot
(ph p b	f p v	voice (noise)	+noi & +voi, glot	lab)
(ph b v	p b f	plosive (voice)	lab, glot	–plos & –voi)

5.7. Typology of inventories

From Figure 14, we see that the inventories { b, f, v } and { ph, f, v } only have arrows that point away from them. This means that our model predicts that these inventories do not exist. The couple of attested inventories reported in Table 6 may have come about by uncommon forces not considered here or may have been incorrectly reported (as suggested by Maddieson 1984:27); it may also be no coincidence that the single reported instance of "{ b, f, v }" has a labial stop (most easy to voice), whereas the single reported instance of "{ ph, f, v }", which in this case simply means { kh, x, ɣ }, has a dorsal stop (least easy to voice).

According to Table 6, the two inventories { ph, b, v } and { ph, p, v } appear to be as uncommon as { b, f, v } and { ph, f, v }. However, their rarity does not require a similar explanation, since it seems simply to be due to the combination of the

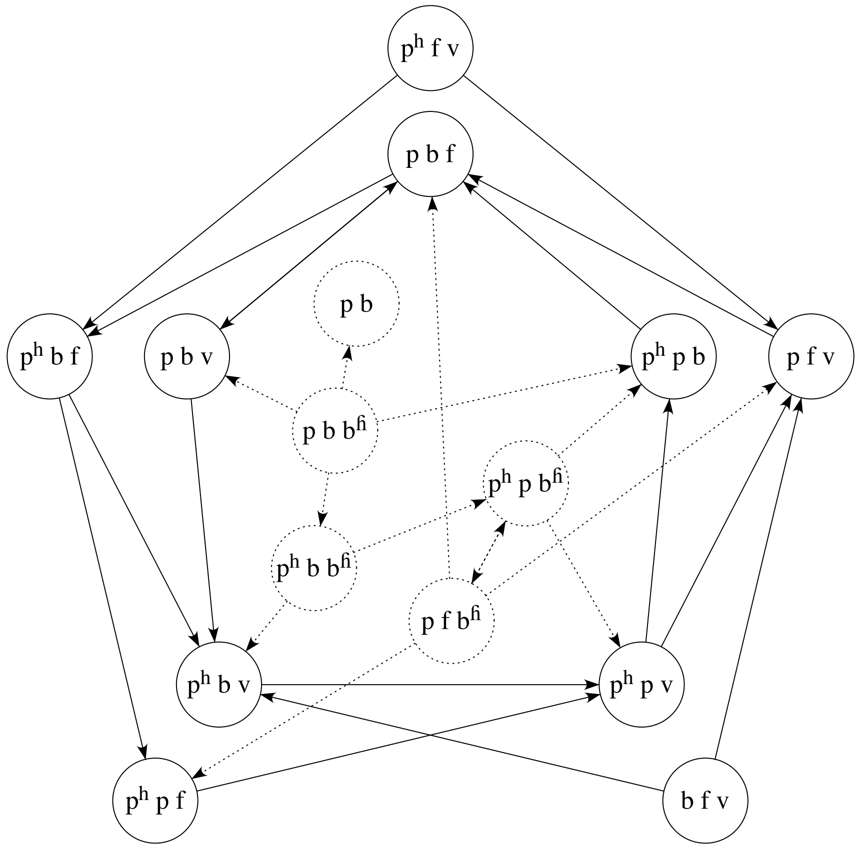

Figure 14. Preferred changes in obstruent inventories.

relative, though not extreme, rarity of the voiced fricative in inventories with a single fricative (compare { p, b, f } with { p, b, v }) and the relative rarity of aspiration in inventories with two plosives (compare { p, b, f } with { ph, b, f }).

In all, we see that our model handles the typological facts pretty well, without having to invoke a separate concept of markedness.

5.8. Attested changes

Most of the arrows in Figure 14 coincide with attested or reconstructed changes in the obstruent systems of the Indo-European languages. Especially the Germanic consonant shifts have traditionally been recognized as cyclic. I will use the development of the words *father*, *two*, and *brother* to illustrate the changes.

Proto-Indo-European has traditionally been regarded as having a { p, b, bɦ }-type inventory for four places of articulation, with a possible gap at the labial |b|

(Brugmann & Delbrück 1897). I will here assume the correctness of this view (for two different theories involving ejective plosives, see Gamkrelidze & Ivanov 1973 and Beekes 1990; for detailed criticisms on these glottalic theories, see Hayward 1989 and Garrett 1991). Thus, Sanskrit had the forms *pitáː*, *duvá-*, and *bɦráːtaː*, as shown in the center of the figure. From Table 4 and Figures 3, 4, 6, 8, and 10 we can derive the possible changes in inventories that contain [bɦ]. The list in Table 14 mentions those that involve inventories that are thinkable intermediate stages for the Indo-European languages (the remaining inventories would only have arrows pointing away from them); the term "brevoi" in the table stands for "glot < [bɦ]".

Table 14. Preferred changes in some inventories of three labial obstruents including |bɦ|

From:	**To:**	**Features:**	**In favor:**	**Against:**
ph p bɦ	ph p b	voice (noise)	+voi, brevoi	lab
ph b bɦ	ph p bɦ	noise (voice)	lab lax	
p b bɦ	ph b bɦ	voice (noise)	–voi lab	glot
p b bɦ	p b ph	noise (voice)	+noi lab brevoi	
p f bɦ	p f b	voice (noise)	+voi brevoi	lab
p f bɦ	p f v	voice (plosive)	+voi brevoi	lab
ph p bɦ	ph p v	voice (noise)	+voi brevoi	lab
p b bɦ	p b v	noise (voice)	+noi brevoi	lab
p f bɦ	p f v	voice (noise)	+voi brevoi	lab
p f bɦ	p f b	voice (plosive)	+voi brevoi	lab
ph p bɦ	f p bɦ	voice (noise)	+noi glot	lab
p f bɦ	p ph bɦ	noise (voice)	–voi lab	glot
ph b bɦ	ph b v	voice (noise)	+noi brevoi	lab
ph b bɦ	ph b v	noise (voice)	+voi brevoi	lab
p f bɦ	p f ph	plosive (noise)	+noi lab brevoi	
p f bɦ	p f ph	noise (plosive)	+plos lab brevoi	

Most of the Indo-European branches did various things about the apparently awkward segment |bɦ|. All of these are shown in Figure 14.

Slavic. In Slavic, |bɦ| merged with |b|, giving the inventory { p, b }.

Greek. Greek turned |bɦ| into the more common |ph|, giving { ph, p, b }: *patéːr*, *dúo*, *phráːteːr*. Table 13 allows a direct two-step change from { ph, p, b } to { p, f, v }. It has been attested in Greek: *patéːr*, *ðýo*, *fráːteːr*. In our model, there is no short single-step route, so we must predict that the two spirantization processes occurred simultaneously. According to Sihler (1995), they should both be dated around the first century A.D.

Latin. Latin probably followed the same route as Greek (Sihler 1995), at least in initial position (Stuart-Smith 1995), subsequently spirantizing the aspirates, thus ending up in { p, b, f }: *pater*, *duo*, *fraːter*.

Early Germanic. The Germanic languages did something very different. The traditional view involves the changes p > ph, bɦ > β (= v), b > p, and ph > f, which could have occurred in any non-neutralizing order. The figure allows four of those orders, all ending up in the Common Germanic { p, f, v } in the figure (**faðár*, **twai*, **βróːθar*). Stopping of initial voiced fricatives then led to the common { p, b, f } inventory (Gothic *faðar*, *twai*, *broːθar*). It is possible, however, that there was no intermediate state with a voiced fricative, i.e. that |bɦ| turned into |b| directly. The figure again allows two possible orders (the method of finding possible sound changes developed in this chapter does not decide between the many proposals found in the literature). These changes caused Germanic to end up in the same inventory as Latin, but with different assignments of the phonemes to the words.

Late Germanic. The Germanic languages have not stopped at { p, b, f }. Most of them have aspirated the voiceless plosive, leading to { ph, b, f } in the figure, and in initial stressed position many of these are in various stages of devoicing the voiced plosive and aspirating the voiceless (English *faːðə*, *tʰʉː*, *b̥ɹʌðə*; High Alemannic *fatːər*, *tswœi*, *b̥ruəd̥ər*; Danish *faː*, *thoˀ*, *pʕoːɒ*; Icelandic *faːðir*, *thʋeir*, *prouðir*), thus leading to { ph, p, f } in the figure. One can say that Icelandic has already taken seven steps along the arrows in Figure 14. High Alemannic went on to spirantize the dorsal affricate, even in initial position (*χuə* 'cow'). Of course, the whole picture is continually complicated by such things as the creation of a new |v| from |w|, the presence of sibilant alveolars, the loss of dorsal fricatives, and loan phonemes such as |v| in English or |f| in Dutch.

Table 13 allows a direct two-step change from { ph, b, v } to { p, b, f }. Since there is no other short route from { ph, b, v } to { p, b, f }, but there is a short route back from { p, b, f } to { ph, b, v } (via { p, b, v }), we have identified another mini-cycle. Moreover, all of these three changes occur within the same underlying feature structure (/plosive/ primary, /voice/ secondary), which could explain why the four inventories { p~ph, b, f~v } can happily live side by side in various parts of the West-Germanic area (Limburgian *vāːdəʀ*, *twíə*, *bʀóːʀ*; Low Saxon *vaːdər*, *thweː*, *broːr*; Western Dutch *faːdər*, *tʋei*, *bruːr*; Westphalian *faːdər*, *thweː*, *broːr*).

In conclusion, the attested cyclic sound changes in Figure 14 can be explained by a model equivalent to the rucksack optimization scheme with three optimizing principles, namely manner faithfulness, place faithfulness, and articulatory effort.

6. HOW LIKELY IS ETERNAL OPTIMIZATION?

Now that we proved that cyclic optimization is possible, is it also the case that it is *likely*? Is the circularity found in §5 an expected outcome, or is this example just a coincidental atypical case and do most other majority-vote optimizations just lead to a stable optimum from which the language can never recover? To find this out, I did two experiments.

6.1. First experiment: independent optimizing principles

I did the following trial 100 times. All ten possible inventories with three initial segments from { p, b, f, v, ph } were ranked randomly on three independent optimizing principles *a*, *b,* and *c*, which could be short for minimization of effort, maximization of manner contrast, and maximization of place contrast. Although the inventories could still represent { p, b, f } and so on, the real-life connection between the labial obstruents and the optimizing principles has been replaced by a random relation. Figure 15 shows two of the 100 results.

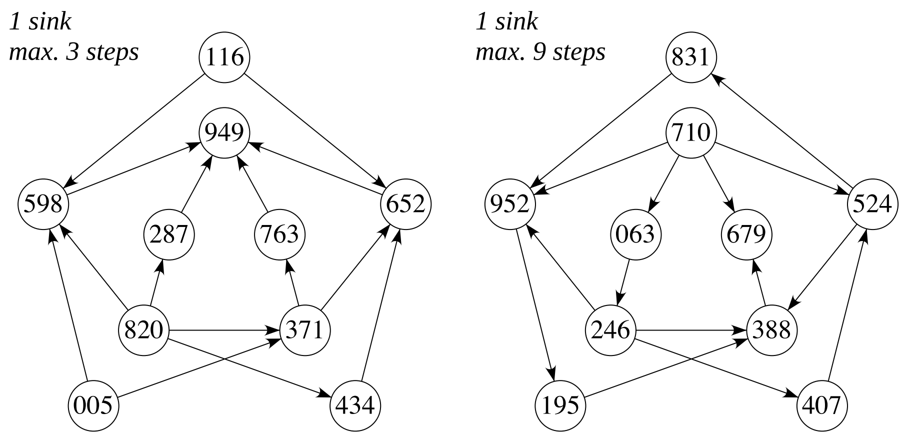

Figure 15. Two absorbing sets of inventories.

As before, the arrows in the graphs represent all 15 preferred single-phoneme changes between the inventories.

The numbers in the graphs are the digit sequences *abc*, from 0 to 9. For instance, the number "598" in the left-hand graph means that this is an inventory in which *a* = 5, *b* = 9, and *c* = 8, e.g. an inventory with intermediate articulatory effort, excellent manner contrast, and very good place contrast. Each of the ten digits (e.g. 5) occurs once as the first digit (598, i.e. *a* = 5), once as the second (652, i.e. *b* = 5), and once as the third (005, i.e. *c* = 5). The arrows show the directions of possible sound changes. For example, there is an arrow from "820" to "371" because 7 is more than 2, and 1 is more than 0, so that two of the three principles (*b* and *c*, e.g. manner contrast and place contrast) favor the "371" inventory over the "820" inventory. There is no arrow from "371" to "820", because only one of the three principles (namely *a*, e.g. minimization of effort) prefers "820" to "371" (because 8 is more than 3); this change, then, is regarded as impossible.

As for the properties regarding cyclicity, there are several possibilities. The two sets in Figure 15 show no cyclicity at all. The left-hand graph has a single *sink* (an absorbing state that allows no subsequent changes to any other state), namely 949,

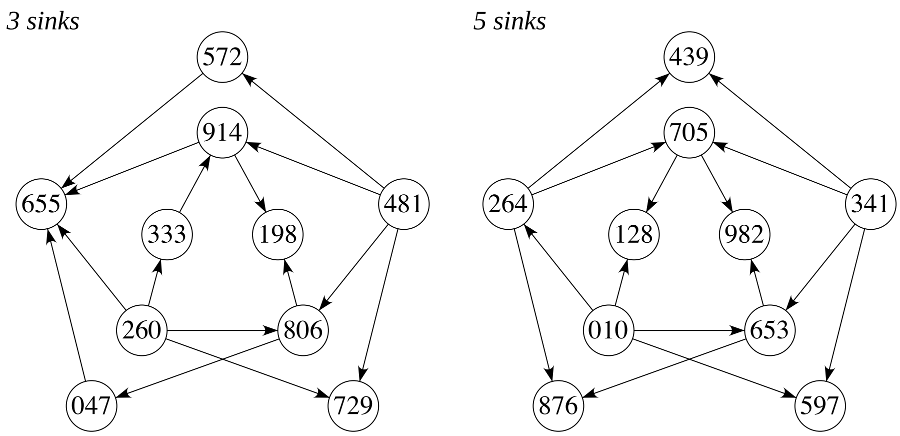

Figure 16. Two sets of inventories with multiple stable states.

which can be reached from any other state (inventory) in at most three steps. This means that regardless of the state (inventory, language) of departure, we will always end up in language "949", i.e. in the language described by the inventory that scores 9, 4, and 9 on the three optimizing principles. The right-hand graph also has a single sink (679), although it may take as many as nine steps to get there, as we can see by following the route starting with 710-063-246-407. Figure 16 shows two graphs with multiple sinks. The left-hand graph has three sinks (655, 198, 729), and the right-hand graph even has five sinks, which means that this graph models a case in which there are five possible stable three-element inventories.

Figure 17 shows examples of cyclic optimization. The left-hand graph shows a 5-cycle (413-926-089-791-802-413) and a 4-cycle (238-089-791-802-238) that is connected to it. If languages have inventories with these optimization principles, they will keep on changing forever. The right-hand graph shows a *leaky* 4-cycle, i.e., every time the language traverses the cycle (780-294-966-078-780), it will have a chance at 294 to leak out of the cycle towards the sink 437, after which sound change will stop (the same for the leak from 780 to 843). Leaky cycles, therefore, show cyclic, but not necessarily eternal, optimization.

Whether leaky cycles are eternal depends on the interpretation of the choices available at the forks. When in state 294, the variation pool may prefer option 966 to 437, simply because it is better in two respects; likewise, when in state 780, the language will prefer 294 to 843. Under this interpretation, the leaky cycle becomes eternal. Eight of the ten possible initial states, then, will lead to this limit cycle, whereas two of the ten initial states will lead to a stable final state. On average, about 50 percent of the initial states in graphs with leaky cycles will end up in an eternal cycle, and the other 50 percent will end up in a sink.

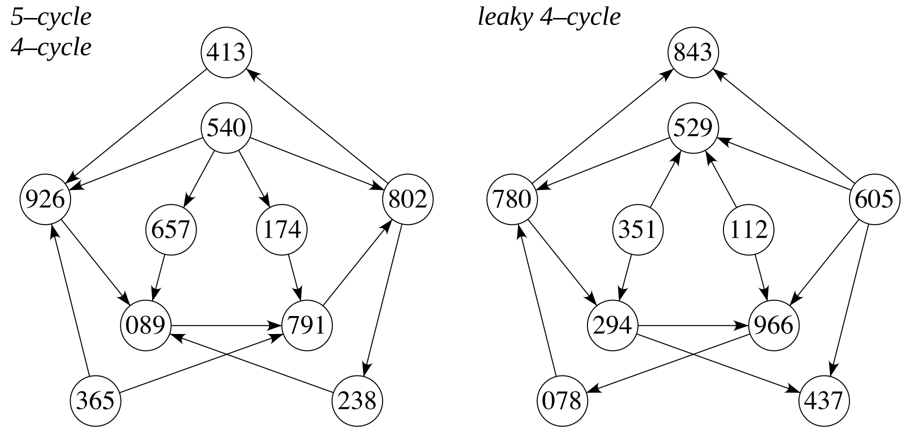

Figure 17. Eternal and finite cyclic optimization.

Unfortunately, not many cyclic graphs were found in this first experiment: in a hundred trials, I found 3 graphs with an eternal cycle, and 6 graphs with leaky cycles.

6.2. Second experiment: dependent optimizing principles

The first experiment was not very realistic: in reality, optimizing principles tend to be dependent on each other, e.g. extra perceptual distinctivity tends to cost additional articulatory effort. So I introduced a dependency between the optimizing principles: *a* and *b* were drawn, independently, from a uniform distribution between –0.5 and +9.5, so that their rounded values could be represented by the digits 0 to 9 with equal probability. The third optimizing principle *c*, however, was chosen to equal 9 minus the average of *a* and *b*. The circles in the graphs in Figure 18 contain rounded values for *abc*.[9] The number 682 in the left-hand graph, for instance, can be explained as follows: the principles *a* and *b* are approximately 6 and 8, respectively, so that their average is about 7; principle *c*, then, is 9 minus this average, i.e. approximately 2.

The left-hand graph in Figure 18 contains five different cycles. These are all connected to each other, and a language may take a different path every time it gets to 544 (though under the single-choice interpretation proposed in §6.1, everything will end up in a single 4-cycle). The right-hand graph contains a 5-cycle (293-474-435-692-952-293) and a sink (942) that is not connected to the cycle. Depending on the initial state, therefore, this graph predicts an eternal circular optimization or a stable inventory.

Fortunately, the second experiment revealed many more cyclic graphs than the first. In a hundred trials, there were 7 graphs with true cycles and 45 graphs with leaky cycles, as summarized in Table 15.

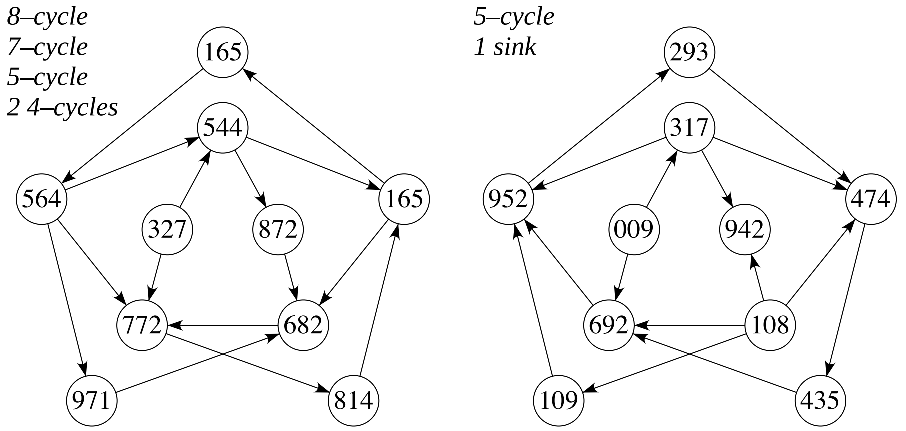

Figure 18. Some eternal optimizations for dependent functional principles.

If functional principles in reality do tend to show trading relationships, as in this second experiment, we can boldly conclude that approximately 50% of all sound inventories are part of a larger set of inventories that includes a cyclic optimization. If we estimate, under the same interpretation as in §6.1, that nearly all of the initial states in the graphs with true cycles lead to an eternal cycle (the right-hand graph of Figure 18 shows one of the very rare exceptions), and that 30 percent of the initial states in the graphs with leaky cycles also end up in an eternal cycle (in half of these graphs, the cycle is eternal, and an average of six initial states will lead to this cycle), then approximately $7 + 0.3 \times 45 = 20$ percent of all initial states in all possible sets of inventories will lead to an eternal loop.

Table 15. Comparing the occurrence of cyclic optimization for two experiments

	cyclic	**leaky**	1 sink	2 sinks	3 sinks	4 sinks	5 sinks
Exp. 1: independent	**3**	**6**	19	35	26	6	5
Exp. 2: dependent	**7**	**45**	20	20	5	3	0

7. CONCLUSION

With the simplest variation scheme that one can think of within the framework of Optimality Theory, sound changes often go on forever, as internal optimization often does not lead to a globally optimal sound system. Thus, optimization by internal functional principles can be a major source of sound change after all. How large the fraction of these changes is in reality, remains to be seen. If all sound change is guided by these internal functional principles, then *all* currently ongoing sound changes are part of a loop, for the simple reason that languages have been around long enough to send all other changes into a sink. External factors, however, will create new initial states, and 80 percent of these will head towards a sink, 20 percent towards a cycle of eternal circular optimization.

University of Amsterdam

8. NOTES

[1] This chapter has profited appreciably from questions and remarks by Benjamin Slade, Eric Holt, Randall Gess, and five anonymous reviewers.

[2] The generalization that 'oral' plosives and nasal stops constitute a natural class in some phonological processes can still be captured by the perceptual feature value /–oral/, which stands for 'no audible oral airflow'.

[3] The term *markedness* serves several purposes. It is also used to refer to the activity of feature values. For instance, [–voice] has been argued to be phonologically inactive (Lombardi 1991). However, Boersma 1998:183 shows that this kind of markedness can be related to the frequency of occurrence of the feature value in the language at hand. This seems unrelated to the role of feature values in determining lexical contrasts. Attempts to reconcile the two types of markedness have been unsuccessful, resulting, for instance, in a divide between the opposing theories of contrastive underspecification (Steriade 1987a) and radical underspecification (Archangeli 1984, 1988).

[4] As a procedure for arriving at an underspecified representation of underlying structure, this is called *feature-tree specification* (Jakobson, Cherry & Halle 1953). It does not share the ambiguities of Steriade's 1987a procedure for contrastive underspecification, in which the only specified feature values are those that can distinguish two segments all on their own; in Table 7, Steriade's procedure would delete both the /–noise/ specification of |b| and the /–voice/ specification of |f|, thereby failing to show how |b| and |f| are contrasted.

[5] Since the Spanish 'spirants' tend to be approximants, this case should really be described in terms of underspecified /plosive/ rather than underspecified /noise/.

[6] Such a perceptual merger may not be impossible, but it is incompatible with the current chapter's intention to restrict itself to inventories of *three* contrasting segments.

[7] This ranking will be different in languages where /v/ has to contrast with /ʋ/. This ranking will also be different for [s] and [z], if [z], as a sibilant, is required to have friction.

[8] The listener's *perception grammar* (Boersma 1998) will probably map an incoming auditory [f] on either of the nearest categories /p/ or /v/.

[9] We see that the rounding hides some information from us: e.g. the arrow from 564 to 544 in the left-hand graph is based on the fact that the 5 in 564 is actually 4.66, and the 5 in 544 is actually 4.95.