

The odds of eternal optimization in OT

Paul Boersma, University of Amsterdam

<http://www.fon.hum.uva.nl/paul/>

December 13, 2000

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 (1997), 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 an optimizing sequence 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.

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.

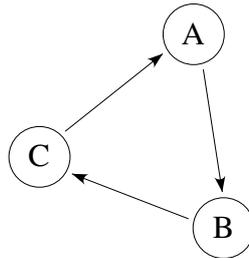
(1) *Three optimizing principles for buying a rucksack*

	<i>volume</i>	<i>weight</i>	<i>price</i>
rucksack A	20 litres	2 kilos	€ 60
rucksack B	30 litres	4 kilos	€ 40
rucksack C	40 litres	3 kilos	€ 90

In our decision which rucksack to buy, we will have to resolve the conflicts between the various optimization principles. 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 consider rucksack A first. We will judge, however, that rucksack B is better than A, because it wins on volume and price, so we will then prefer B. However, rucksack C is better than B regarding volume and weight, so we must prefer C to B. However, we cannot buy rucksack C, because A is better on weight and price. Figure (2) shows how our decision will cycle about in a loop.

(2) *The simplest eternal optimization scheme*



The conclusion must be that eternal optimization is possible, if quality is defined by a majority vote among optimizing principles.

2. Teleological eternal optimization of sound systems

Boersma (1989) applied the above optimization scheme to inventories of three labial obstruents chosen from the set { p, b, f, v, ph } in accented initial position. Two examples of such inventories are { p, b, f } and { b, v, ph }. In total, there are ten possible inventories of this type. The three optimizing principles were “minimize articulatory effort”, “maximize perceptual contrast (i.e. the manner contrast between obstruents)”, and “maximize perceptual salience (i.e. the perceptual contrast between the obstruent and the following vowel)”. Sound change, then, was modelled as follows:

(3) *Teleological sound change*

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

Phonemes are considered *adjacent* if they are adjacent in the sequence p-b-v-f-ph-p. Likewise, two inventories are considered adjacent if they differ in only one pair of adjacent elements, e.g. { p, b, f } is adjacent to { ph, b, f }, { p, b, v }, { p, v, f }, and { p, b, ph }. Table (4) compares several pairs of inventories on the three optimizing principles. For instance, the inventory { ph, b, f } is better than { p, b, f } because it wins on perceptual contrast (the *ph-b* distinction is better than the *p-b* distinction) and on perceptual salience (*pha* is a perceptually more salient sequence than *pa*).

(4) *Eternal optimization in consonant inventories (Boersma 1989)*

	<i>minimize articulatory effort</i>	<i>maximize perceptual contrast</i>	<i>maximize perceptual salience</i>
p b f			
↓	worse	better	better
ph b f			
↓	better	worse	better
ph p f			
↓	better	better	worse
ph p v			
↓	better	worse	better
ph p b			
↓	better	better	worse
f p b			

After five optimizing steps, we are left with the initial inventory, though all consonants have shifted. It is as if Latin (*pater, duo, frater*) becomes Germanic (*father, two, brother*) in five steps.¹

The phonetic details of the optimization will become clear in the Optimality-Theoretic account, which follows in the next section.

3. The underlying blind mechanism: invisible ranking

As summarized in the previous section, Boersma (1989) showed that if the success of a proposed sound change is determined by a majority vote among a number of functional principles, the language may go on changing forever, even if no external factors appear on stage. The drawback of this approach is that the selection step is teleological, i.e. goal-oriented. Finding a blind underlying mechanism to account for the selection step would be more satisfying.

One such blind mechanism is provided by Optimality Theory, in which it seems natural that variation can be described as a result of a set of mutually unranked constraints. 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 (1997) used the following variation-and-selection model for predicting the direction of sound change:

¹ The d-t pair is included for lack of a good b-p pair. Of course, it is likely that Germanic *b* does not come from *f*, but Germanic *b* and Latin *f* stem from a common ancestor commonly known as *bh*.

(5) *Non-teleological sound change*

- a. Start with any inventory and determine its Optimality-Theoretic constraint-ranking grammar.
- b. **Variation:** the workings of many constraints are invisible, so that their mutual ranking is random, i.e. different for every speaker. A lack of contrastiveness then causes one faithfulness constraint to fall. The formerly hidden rankings now become visible, which reveals several new sound systems.
- c. **Non-teleological selection** (reanalysis): from the pool of variation, the next generation chooses the sound system that occurs most often in this pool. This can be seen as a postponed majority decision among the speakers of the language.
- d. Return to step b.

In the following sections, I will discuss what constraints rankings can be regarded as fixed, and what rankings must be language-dependent.

4. Fixed rankings in obstruent systems

According to Prince & Smolensky's (1993) concept of *harmonic ordering*, some Optimality-Theoretic constraint families can be internally ranked in a language-independent way. According to the theory of Functional Phonology (Boersma 1998), the production grammar consists of articulatory constraints and perceptual faithfulness constraints. This theory proposes a set of *local-ranking principles*, according to which these constraints can be ranked by their satisfaction of functional principles. For our set of obstruents, the fixed rankings are listed in (6).

(6) *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 one partly fixed hierarchy of perceptual place faithfulness constraints (§4.2).
- c. Maximization of three perceptual 'manner' contrasts (§4.3):
 - the voicing contrast, giving five fixed hierarchies of voicing faithfulness;
 - the noisiness contrast, giving five fixed faithfulness hierarchies;
 - the plosiveness contrast, giving five fixed faithfulness hierarchies.

All the hierarchies in the following three sections are taken directly from Boersma (1997).

4.1 Fixed hierarchy for articulatory effort

According to the local-ranking principle for articulatory constraints (Boersma 1998: ch. 7), 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. 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 the active glottal spreading gesture must be stronger if the supralaryngeal vocal tract is

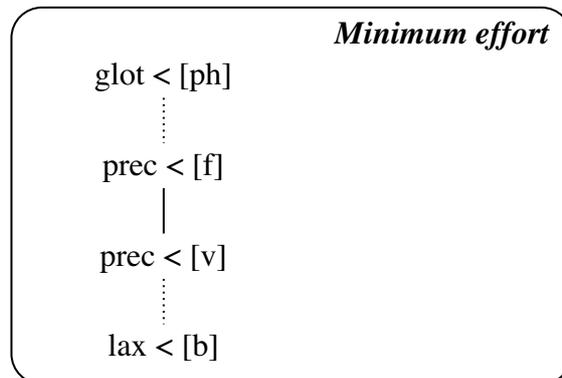
unimpeded (as in the aspiration phase of [pha]) than if the oral and nasal cavities are wholly or nearly sealed off (as during the closure periods of [pa] and [fa], when voicelessness is called for). We can express this as the fixed ranking “glot < [ph]” >> “glot < [f]” >> “glot < [p]”, where “glot < [x]” is an abbreviation for “do not perform a glottal spreading gesture at least as difficult as that required for a typical [x]”.

Likewise, we can posit a 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 pronounced as the approximant [v], 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]”.²

Third, 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, 1997), these fixed rankings were simplified to the hierarchy in (7).

(7) *Hierarchy of articulatory constraints*



In this picture, the universal ranking of the two precision constraints is given by the solid line. According to the local-ranking principle, the two 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.

4.2 Fixed hierarchy for faithfulness of perceptual place

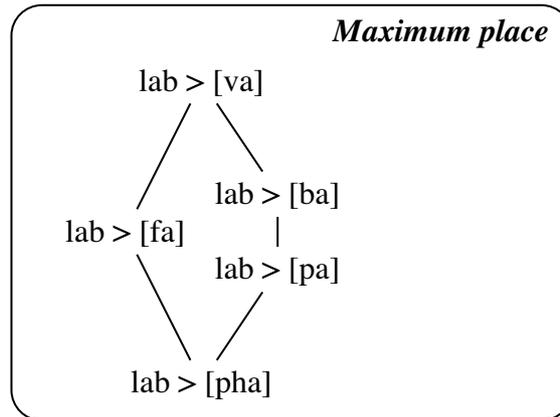
For faithfulness constraints, I will consider all universal local hierarchies, and posit no globally fixed rankings (in contrast with the global articulatory ranking of §4.1).

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. An example of a labiality faithfulness constraint can be “lab > [ba]”, which is shorthand for “the acoustic cues for labiality

² This ranking will be different in languages where /v/ has to contrast with /v/. This ranking will also be different for [s] and [z], if [z], as a sibilant, is required to have friction.

should be better than 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. According to the local-ranking principle for faithfulness constraints (Boersma 1998: ch. 9), these constraints are ranked higher if their violation would cause more confusion. Since having [v]-like place cues causes more perceptual confusion than having [b]-like place cues, the constraint “lab > [va]” must outrank “lab > [ba]”. The whole relevant hierarchy is shown in (8).

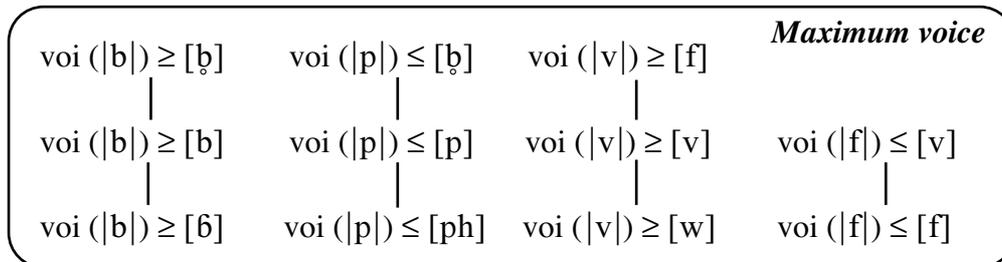
(8) *One fixed hierarchy for place faithfulness*



4.3 Fifteen fixed hierarchies for perceptual ‘manner’ faithfulness

Analogous hierarchies can be posited for manner features. A segment specified underlyingly for [+voice] should be pronounced with as many voicing cues as possible if it has to contrast with a voiceless segment. 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|) ≥ [ɓ]” >> “voi (|b|) ≥ [b]” >> “voi (|b|) ≥ [b̥]”. An analogous hierarchy can be posited for the voiced labial fricative and for the three voiceless segments. Figure (9) shows four of the five universal hierarchies.

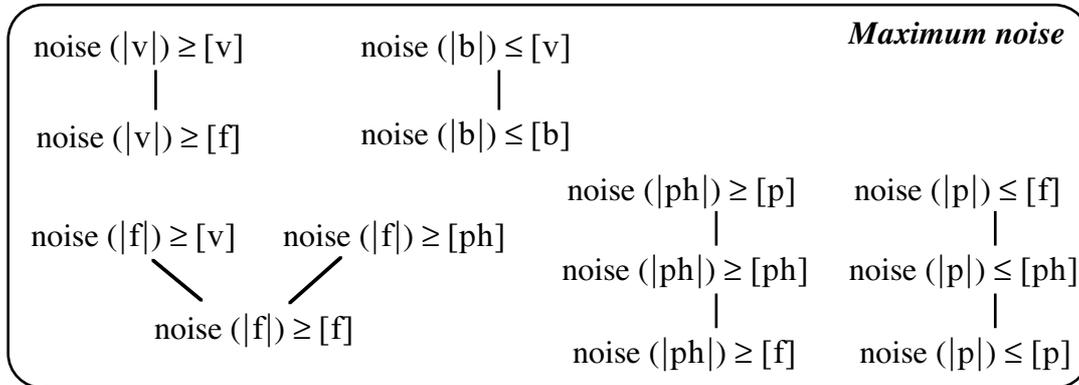
(9) *Fixed hierarchies for voicing faithfulness*



The hierarchy for |ph|, not shown here, is identical to that for |p|. The solid lines depict the fixed rankings, and the five hierarchies are freely ranked with respect to each other, e.g., “voi (|p|) ≤ [p̥]” could outrank “voi (|v|) ≥ [f]” in some languages.

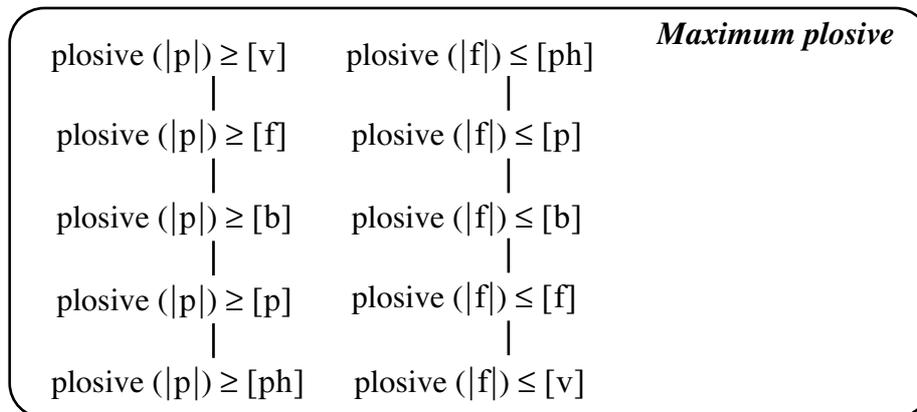
Segments specified underlyingly for [+noise], i.e. fricatives and |ph|, should be pronounced with as many noisiness cues as possible. The voiceless fricative [f] will be noisier than the voiced fricative [v] and the aspirated plosive [ph], which leads to the hierarchy in the bottom left corner of figure (10). The other four hierarchies are constructed in the same way.

(10) *Fixed hierarchies for noise faithfulness*



Finally, the five segments divide into three plosives and two fricatives. If we assume that the best plosive is a voiceless plosive, we get the plosiveness or continuancy hierarchies in (11).

(11) *Fixed hierarchies for plosive faithfulness*



The hierarchies for |b| and |ph| are identical to the one for |p|, and the hierarchy for |v| is identical to that for |f|.

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 §3 and given the fixed rankings of §4. I will generalize this example to the other possible 11 (or 13) changes within the set of three-obstruent inventories, 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

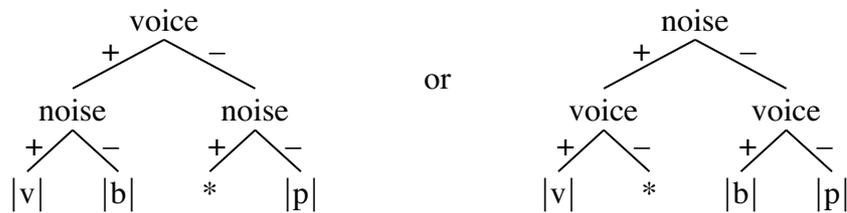
There are three ways to describe a language with a { p, b, v } inventory without variation. The first way to describe this inventory is as a full specification of the voicing and noise features, as in (12).

(12) *Full specification*

	p	b	v
[voice]	–	+	+
[noise]	–	–	+

The second way is with a feature-tree specification (Jakobson, Cherry & Halle 1953), as in (13).

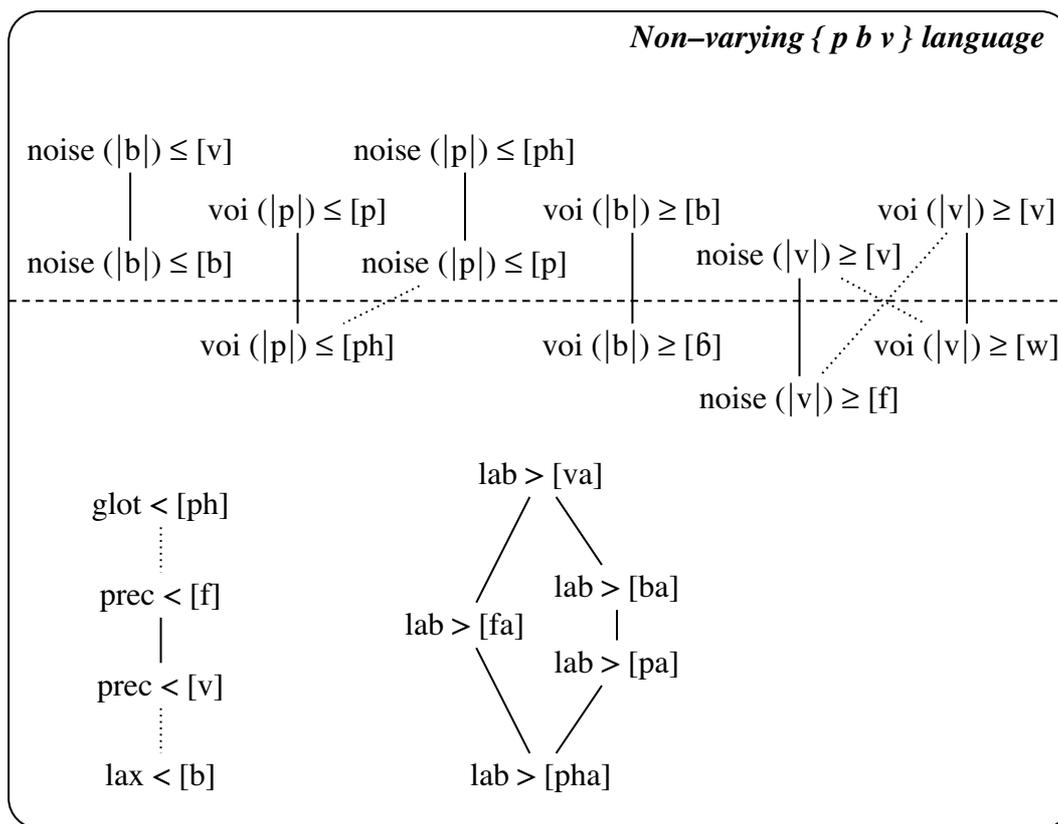
(13) *Two possible feature trees*



We see that there are two possible feature trees, one that opposes the two voiced segments to the single voiceless segments, and one that opposes the plosives to the fricative. Both feature trees have a gap at [f].

The third way to describe this language is with an Optimality-Theoretic constraint grammar, as in (14).

(14) *Constraint grammar*



In this grammar, we see seven of the 17 fixed hierarchies of §4. The constraints above the dashed line are said to be in the *first stratum*, since all of them are undominated. 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).

I will give some examples of the realization of underlying segments. Tableau (15) shows that an underlying $|p|$, i.e. “voiceless non-noisy” according to (13), is pronounced as $[p]$.

(15) *The pronunciation of the voiceless non-noisy obstruent*

$ p $ i.e. [-voice, -noise]	noise ($ p $) ≤ [ph]	noise ($ p $) ≤ [p]	voi ($ p $) ≤ [p]	voi ($ p $) ≤ [ph]	lab > [pa]
[ph]		*!			
[p]				*	*
[b]			*!	*	*
[v]	*!	*	*	*	*
[f]	*!	*			*

Note that “noise ($|p|$) \leq $[p]$ ”, i.e. a constraint above the line in (14), crucially outranks “voi ($|p|$) \leq $[ph]$ ”, a constraint below the line. Otherwise, $[ph]$ would have been the winner. Several crucial rankings like this one show up as dotted lines in (14).

In (16), we see that an underlying $|b|$, i.e. “voiced non-noisy”, is pronounced as $[\beta]$.

(16) *The pronunciation of the voiced non-noisy obstruent*

$ b $ i.e. [+voice, -noise]	noise ($ b $) \leq $[v]$	noise ($ b $) \leq $[b]$	voi ($ b $) \geq $[b]$	voi ($ b $) \geq $[\beta]$	lax $<$ $[b]$
$[ph]$		*!	*	*	
$[p]$			*!	*	
$[b]$				*!	*
 $[\beta]$					
$[v]$		*!			
$[f]$	*!	*	*	*	

This is not really what we want. Apparently, we have to assume an articulatory constraint against the implementation of an implosive, e.g. *GESTURE (lowering larynx), in the first stratum. This will yield the desired $[b]$ as the winning candidate.

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

The full specification (12) seems redundant. Surely a language can change at least one of the six feature values without destroying comprehension. As a criterion for free variation, therefore, we could say that segments are allowed to vary freely as long as the listener can easily reconstruct the underlying form.

We can describe the variation in three ways again. The first way is by specification of features. According to Steriade’s (1987) algorithm for contrastive underspecification, two feature values can be deleted from table (12), giving table (17), in which the deleted feature values have been put between parentheses.

(17) *Contrastive underspecification theory*

	p	b	v
[voice]	–	+	(+)
[noise]	(–)	–	+

As a scheme for variation, this theory is too weak. If voicing and noise are the only contrasts, then (17) says that $|p|$ is allowed to be pronounced with noise, i.e. as $[f]$, and $|v|$ is allowed to be pronounced without voice, i.e. also as $[f]$. Now that $[f]$ is a variant of $|p|$ as well as $|v|$, the underlying segment is no longer reconstructable from the surface form. This causes an amount of perceptual merger that is incompatible with my intention to restrict myself to inventories of three contrasting segments. Therefore, we could say that

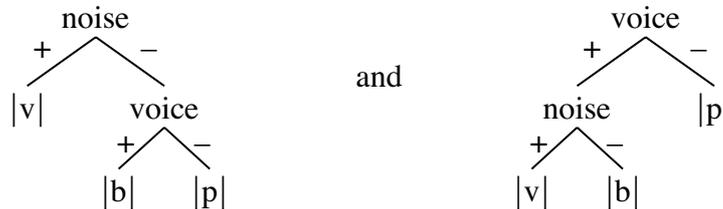
only one of the two feature values may change, i.e., we have either of the two possibilities in (18).

(18) *Allowed underspecifications*

	p	b	v	and		p	b	v
[voice]	–	+	(+)		[voice]	–	+	+
[noise]	–	–	+		[noise]	(–)	–	+

In the left table, [f] is a variant of |v|, whereas in the right table, [f] is a variant of |p|. Feature-tree underspecification (Jakobson, Cherry & Halle 1953) does not have the disadvantage of the contrastive underspecification of (17). There are only two possibilities, shown in (19).

(19) *Feature-tree underspecification*



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 in the left table of (18), and the feature tree is the left-hand tree in (19). Both of these underspecifications, however, are 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 (20) 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.

(20) *The pronunciation of the voiced noisy obstruent*

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

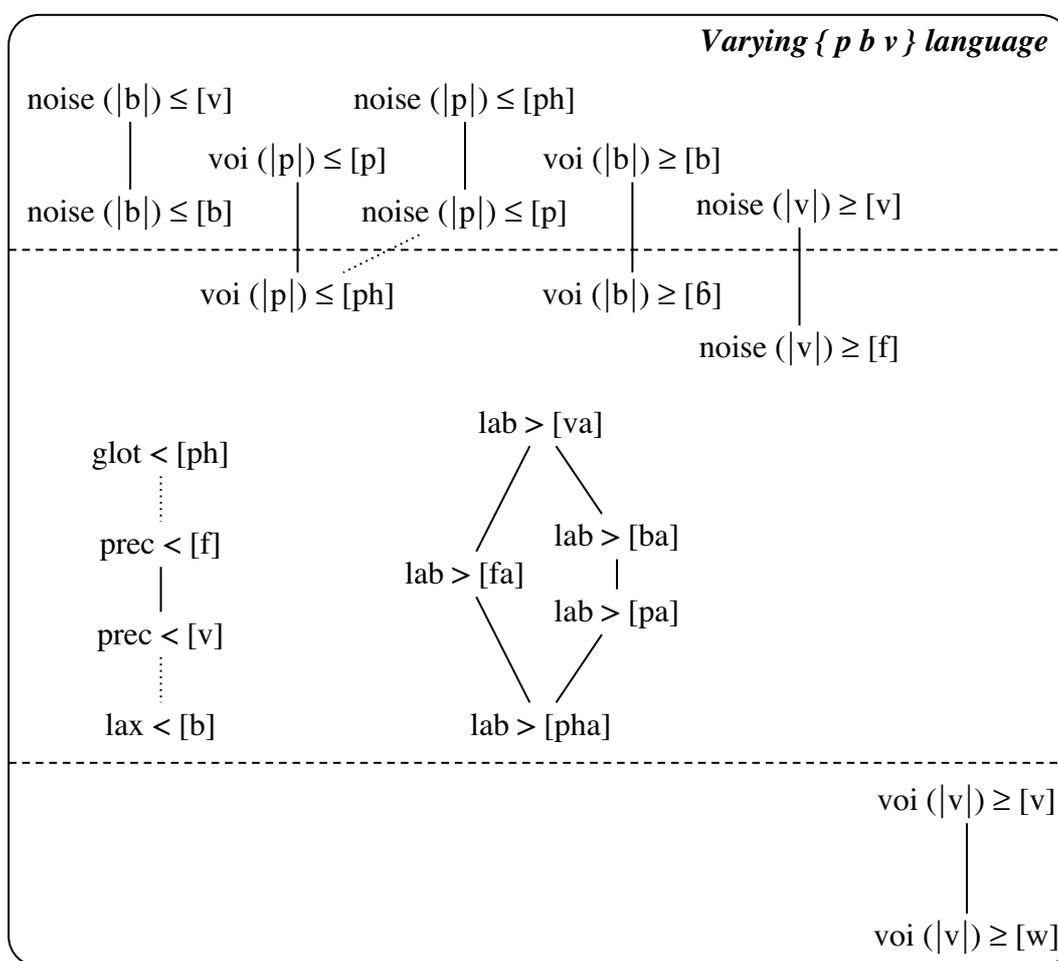
In the same manner, however, post-obstruent |v| will be devoiced, as shown in (21).

(21) *The pronunciation of the voiced noisy obstruent*

$ atva $ i.e. [+voice, +noise]	*[voiced fricative / obstruent _]	voi ($ v $) $\geq [v]$	lab > [va]	noise ($ v $) $\geq [f]$
 [atva]	*!		*	*
[atfa]		*		

But according to the principle of perceptual recoverability, the variation between [v] and [f] realizations could be completely free, i.e. “voi ($|v|$) $\geq [v]$ ” could be ranked very low, perhaps in a third stratum. This extreme version of Dutch is shown in figure (20).

(22) *Constraint-ranking grammar*



Now that the [+voice] specification for |v| has fallen down 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 (14). There are three relevant constraints here: “noise ($|v|$) $\geq [f]$ ”, “prec < [f]”, and “lab > [va]”.

These three constraints are ranked in an unpredictable order in the pool of between-speaker variation. If “noise ($|v|$) $\geq [f]$ ” happens to be ranked on top of these three, the noisiness contrast of |v| with respect to |b| and |p| will be enhanced, as shown in (23).

(23) *Enhancement of noisiness contrast*

pabava	noise (v) ≥ [f]	prec < [f]	lab > [va]	voi (v) ≥ [v]
[pabava]	*!		*	
☞ [pabafa]		*		*

If “prec < [f]” is on top, the input will surface faithfully, as shown in (24).

(24) *Minimizing precision*

pabava	prec < [f]	noise (v) ≥ [f]	lab > [va]	voi (v) ≥ [v]
☞ [pabava]		*	*	
[pabafa]	*!			*

And if “lab > [va]” is on top, the place contrast of |v| with respect to other fricatives such as |ð| and |ɣ| will be enhanced, as shown in (25).

(25) *Enhancement of place contrast*

pabava	lab > [va]	noise (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 devolve an underlying |v|.

This section showed 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

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 |pabava| to |pabafa| in two generations.

One would think that the reanalysis step does not lead to a change in the surface forms. After all, the voiceless specification constraint “voi (|f|) ≤ [f]” will be ranked in the bottom stratum, 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 surface inventory will have changed from a non-variable { p, b, v } to a non-variable { p, b, f }. 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. That is, the changes in the rankings of these constraints have no direction; they go up and down the hierarchy. The result, 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 Predicted possible sound changes

The { p, b, v } to { p, b, f } change is just one of the 14 changes predicted by the OT variation-and-selection scheme. The others are listed in table (26).

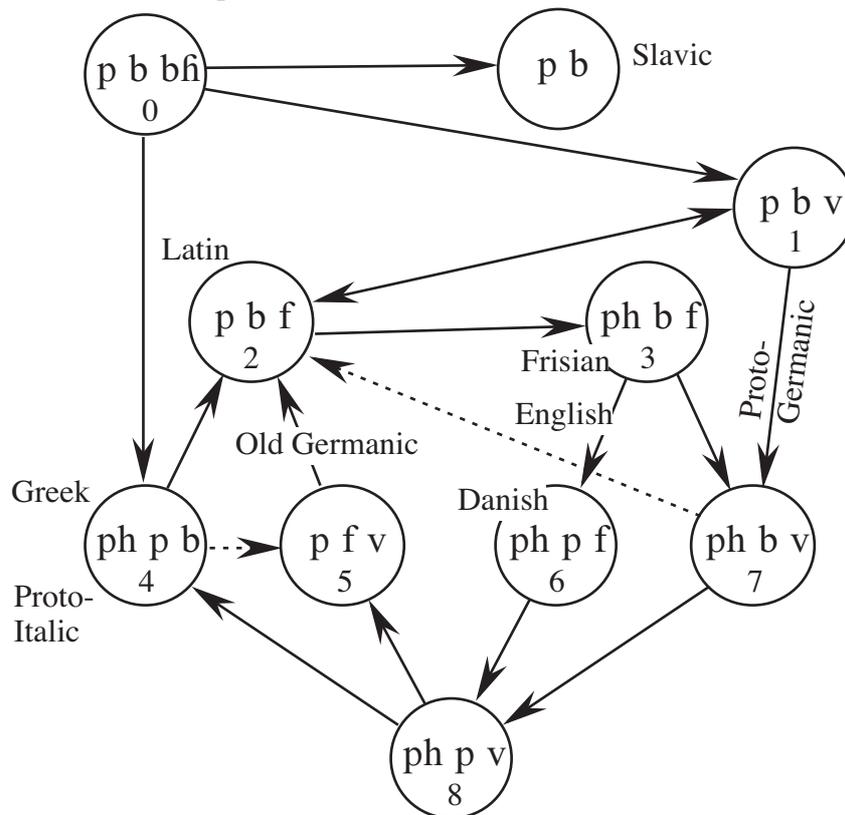
(26) *Possible changes in inventories of three labial obstruents*

From:	To:	Feature tree:	In favour:	Against:
p b v	p b f	noi voi	+noi, lab	prec-f
p b v	ph b v	plos voi	-voi, lab	glot
p b f	ph b f	plos voi	-voi, lab	glot
p b f	p b v	plos voi	-plos, prec-f	lab
p v f	p b f	voi noi	prec-v, lab	+voi
ph p b	f p b	voi noi, noi voi	+noi, glot	lab
(ph p b	f p v	voi noi	+noi & +voi, glot	lab)
ph p v	ph p b	voi noi	prec-v, lab	+voi
ph p v	f p v	voi noi	+noi, glot	lab
(ph b v	p b f	plos voi	lab, glot	-plos & -voi)
ph b v	ph p v	plos noi	lab, lax	
ph b f	ph b v	plos voi	-plos, prec-f	lab
ph b f	ph p f	plos noi	lab, lax	
ph p f	ph p v	plos noi	-plos, prec-f	lab

The example of §5.1–3 is summarized in the first row of (26): 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 favour of the change were noise faithfulness and labiality faithfulness, and the constraint that voted against the change was minimization of precision. To include |ph| in the inventory, table (26) includes the third feature, i.e. [plosive]. The two changes between parentheses are changes of two features at the same time (the change from { ph, p, b } to { f, p, v } may have occurred in Greek around the year zero). Figure (27) shows all of these 14 possible changes, plus three changes emanating from an alleged { p, b, bfi } inventory. The figure shows some of the languages that can be associated with these changes. The history of English, for instance, can be regarded as starting in inventory { p, b, bfi } (Proto-Indo-European), going by aspiration of PIE /p/ and frication of /bfi/ to a Proto-Germanic { ph, b, β } system (= 7), then by spirantization and devoicing to a Common Germanic { f, p, β } system (= 5), then by stopping and another aspiration to { f, ph, b } (= 3), which is underway with another devoicing to { f, ph, p } (= 6), at least in prevocalic stressed position (ignoring late developments like the loan phoneme |v|).

(27) *Circular changes in obstruent inventories*

Proto-Indo-European



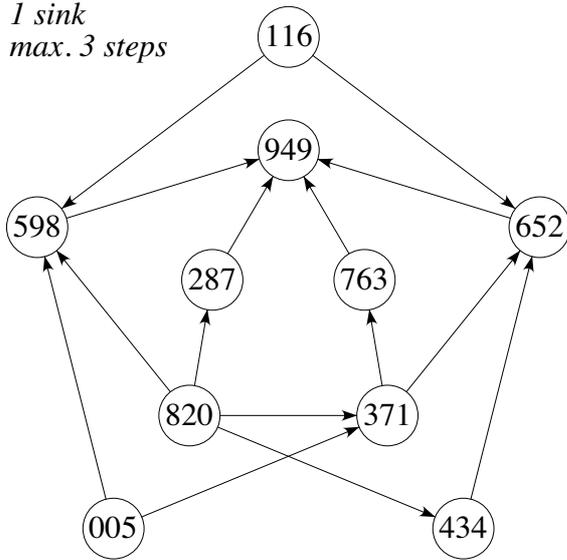
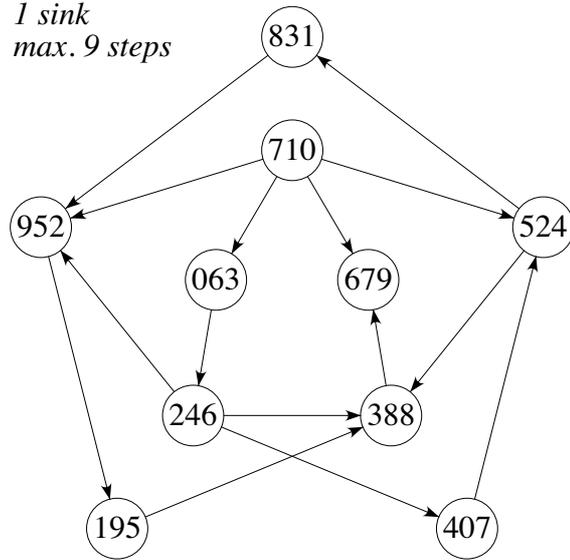
In conclusion, the circular optimization in (27) is equivalent to the rucksack optimization scheme with three optimizing principles, namely manner faithfulness, place faithfulness, and articulatory effort. It proves that cyclically optimizing sound changes are possible.

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 segments from $\{ p, b, f, v, ph \}$ are ranked randomly on three independent optimizing principles a, b , and c . Figure (28) shows two of the 100 results.

(28) *Two absorbing sets of inventories**1 sink
max. 3 steps**1 sink
max. 9 steps*

The connections in the graphs represent all 15 possible 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 $c = 8$. 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) favour the “371” inventory over the “820” inventory.

As for the properties regarding cyclicity, there are several possibilities. The two sets in (28) show no cyclicity at all. The left-hand graph has a single *sink* (absorbing state), namely 949, 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 (29) 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.

(29) *Two sets of inventories with multiple stable states*

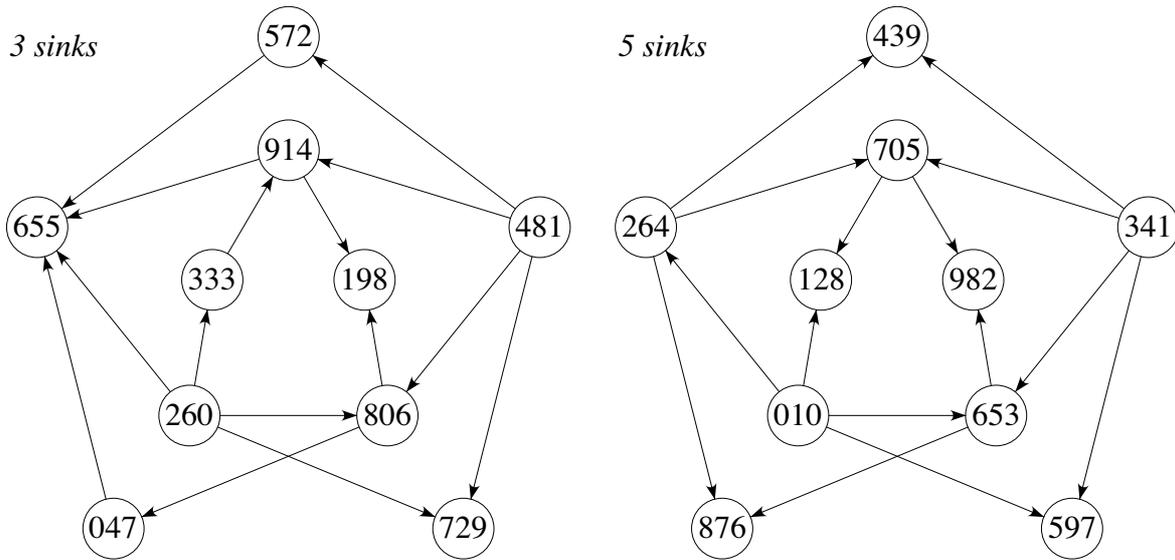
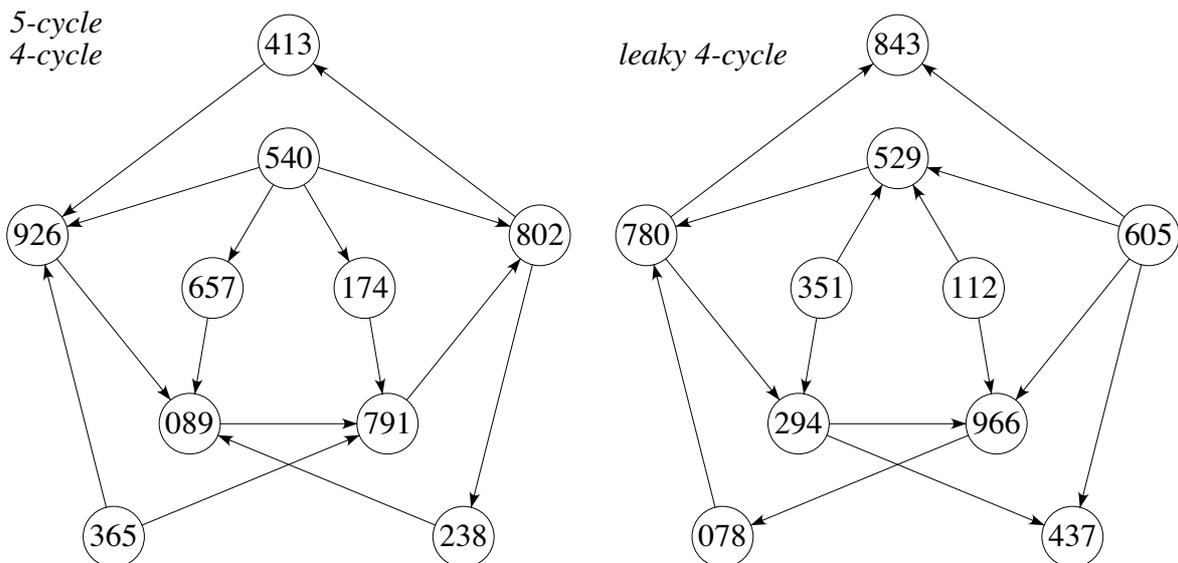


Figure (30) shows examples of cyclic optimization. The left-hand graph shows a 5-cycle (413-926-089-791-802-413) and a 4-cycle (238-089-791-802-238) that is connected to it. If languages have inventories with these optimization principles, they will keep on changing forever. The right-hand graph shows 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.

(30) *Eternal and finite cyclic 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

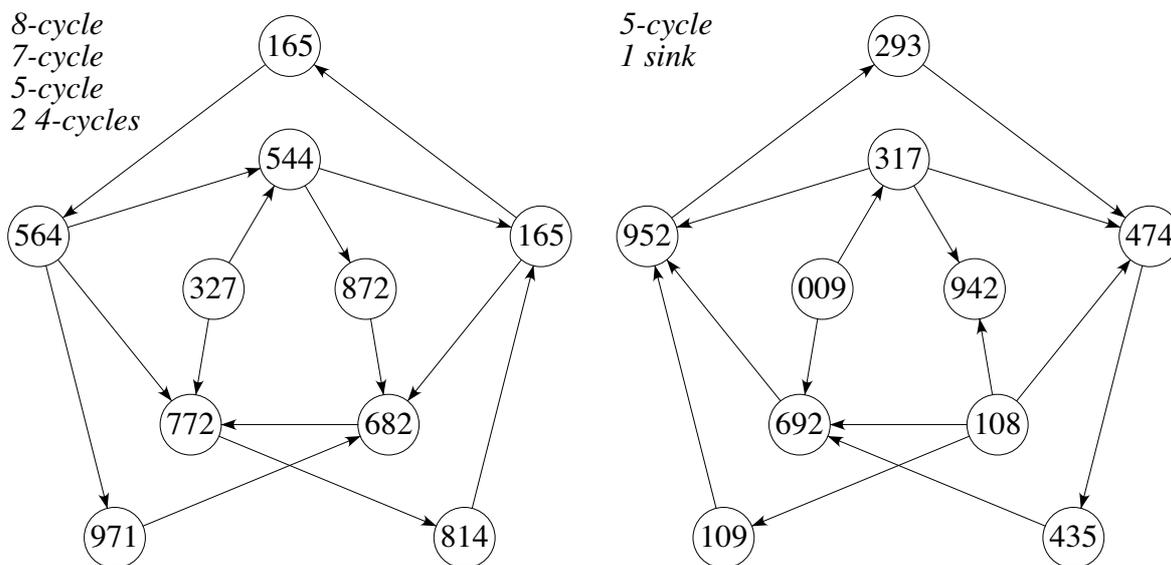
possible initial states, then, will lead to this limit cycle, whereas two of the ten initial states will lead to a stable final state. On average, about 50 percent of the initial states in graphs with leaky cycles will end up in an eternal cycle, and the other 50 percent will end up in a sink.

Unfortunately, not many cyclic graphs were found in this first experiment: in a hundred trials, I found 3 graphs with an eternal cycle, and 7 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 graphs in (31) show the rounded values for abc .³ 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.

(31) *Some eternal optimizations for dependent functional principles*



The left-hand graph in (31) contains five different cycles. These are all connected to each other, and a language may take a different path each time it gets to 544 (under the single-choice interpretation proposed in §6.1, everything ends 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.

³ 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.

Fortunately, the second experiment revealed many more cyclic graphs than the first. In a hundred trials, there were 6 graphs with true cycles and 45 graphs with leaky cycles, as summarized in table (32).

(32) *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

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 (31) shows one of the very rare exceptions, with only 9 out of 10 initial states leading to an eternal cycle), 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.

7. Conclusion

With the simplest OT variation scheme, 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.

References

- Boersma, Paul (1989). 'Modelling the distribution of consonant inventories by taking a functionalist approach to sound change.' *Proceedings of the Institute of Phonetic Sciences* 13: 107–123. University of Amsterdam.
- Boersma, Paul (1997). 'Sound change in Functional Phonology.' Rutgers Optimality Archive 237, <http://ruccs.rutgers.edu/roa.html>. [Identical to chapter 17 of Boersma 1998]
- Boersma, Paul (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Doctoral thesis, Univ. of Amsterdam. The Hague: Holland Academic Graphics.
- Jakobson, R., E.C. Cherry & M. Halle (1953). 'Toward the logical description of languages in their phonemic aspect.' *Language* 29: 34–46.
- Prince, Alan & Paul Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report TR-2, Rutgers University Center for Cognitive Science.
- Steriade, Donca (1987). 'Redundant values.' In A. Bosch, B. Need & E. Schiller (eds.) *CLS 23: Papers from the Parasession on Autosegmental and Metrical Phonology*. Chicago Linguistic Society. 339–362.