A simple account of chain shifts and saltations in Optimality Theory

Charles Reiss
Concordia University, Montréal

Abstract

Contrary to universally held assumptions, classic OT constraint interaction, without conjoined constraints, distance faithfulness or P-Maps, can generate chain shift and saltation patterns. These new results are derived by constructing OT grammars with the required properties. The key is to derive the relevant alternating segments from underspecified segments, thus reconceptualizing these patterns as involving mappings among elements, some of which may possibly never appear in output forms. It is now apparent that the presumed failure of classic OT to generate chainshifts cannot play a role in evaluating the merits of classic OT *vis-à-vis* other models of phonological computation.

1 Introduction

It has been universally accepted since the advent of Optimality Theory (OT) that "chain shifts cannot be analyzed under standard OT" (Kager, 1999, 394). The logic behind this view was also extended to saltations. As White (2017) puts it "saltations cannot be derived in 'classical OT' " because of "the excessive nature of the change involved in such alternations". These views led to suggestions for innovations of the original OT model, conveniently discussed with regard to chain shifts by McCarthy (2002, 162), such as local constraint conjunction (Kirchner, 1996; Smolensky, 1995), scalar faithfulness (Gnanadesikan, 1997) or faithfulness to input context. Saltations have also driven OT practitioners to propose enhancements to the framework, most prominently the version proposed by White (White, 2017; Hayes and White,

2015) involving the P-Map (Steriade, 2008, 2001) and *MAP constraints (Zuraw, 2007). I'll clarify what these terms, 'chain shift' and 'saltation' mean below, because it turns out that under the definitions I propose, the accepted wisdom is false: these patterns *can* be analyzed under standard OT. This is a new result.

The chain shift and saltation problems in the OT literature are presented in terms of mappings among fully specified segments, mirroring the traditional usage of these terms in both diachronic phonology and earlier generative models. The discussion here does not solve the chain shift and saltation problems as presented in the OT literature, using such fully specified segments, but instead re-conceptualizes the problem. If we consider the two terms as referring to relationships among the segments that surface in morphologically related forms—forms containing the same lexical item—and we make use of underspecified inputs, then deriving the two phenomena becomes simple, using machinery that has long been available, and in particular was available in the days of standard or "classical" OT. Despite the fact that the problem is re-conceptualized here, the important result is that the the surface patterns associated with chain shifts and saltations can actually be derived with classical machinery, with the obvious implication that enhancements to that machinery need to be re-examined.

This paper is not an evaluation of those enhancements, the other approaches to the patterns in question, such as the use of local conjunction of constraints. There is no claim that the solution offered here is 'better' than one involving local conjunction, say. That question is beyond the scope of our concerns, and those other approaches will not be discussed further. It should be obvious that if there are other compelling arguments for, say, constraint conjunction, then the usefulness of the results in this paper would be diminished. My goal is to neither over- nor understate the results I claim. The paper merely provides construction proofs of classical OT solutions to generating the relevant surface patterns. As discussed below, this humble result should play some role in any sober discussion of theory comparison.

Two readers of previous versions of this work have insisted that it is necessary to consider the role of the OT Principle of Richness of the Base (RotB) in the analyses I provide. I do not think that this paper is the place for a full discussion of RotB in light of the important role that the principle has played in the OT literature. However, I will provide a brief indication of problems I see with RotB in general. If my objections to the status of RotB are valid, then the relevance of RotB to the analysis offered here, indeed to any analysis, is compromised.

2 Underspecification

Because the analyses offered here involve underspecified inputs, some discussion of underspecification is in order. To start, it is necessary to adopt a theory of representations. I assume here that features are binary, in the sense that features occur in representations as part of an ordered pair consisting of a value from the set $\{+, -\}$ followed by a feature F_i from a universal feature set $\mathbf{F}$. This binarity assumption is widespread in the OT literature. For present purposes, we'll assume that such pairs, +VOICED or −ROUND, to use familiar examples, are members of the set that constitutes a segment (along with other information such as associated timing slots). In other words, I am not adopting a privative feature system. Underspecification just refers to the possibility that a segment may lack one or more members that contain some features, F_k in $\mathbf{F}$. That is, segments may be 'incomplete' with respect to $\mathbf{F}$. A segment may contain $+F_i$, or $-F_i$, or neither.

Now, there is nothing inherent in "classical OT" that prohibits underspecification,. Furthermore, the literature is quite clear that logical issues such as the possibility of classic OT's capacity to handle chain shifts is orthogonal to the particular representational system adopted: "Optimality Theory is a theory of constraint interaction, not of representations. We want our deductions about OT to hold even if the theory of representations changes" (Moreton, 1999, 142).

Nonetheless, to forestall objections to the demonstrations provided below, I now offer three arguments in favor of allowing underspecification in theories of phonological representation. A first argument for the existence of underspecification in the segments occurring in lexical entries (which supply the segments appearing in the *input* to the phonology) is provided by Keating (1988); Choi (1992) and others who suggest that underspecification exists even in the *output* of the phonology. While some discussions assume that underlyingly absent features are provided by feature-filling processes, this body of work argues that underspecification can persevere into the output in ways that are supported by phonetic analysis. This phenomenon is referred to as surface, or phonetic, underspecification.

A second argument for underspecification is that there is no good argument against it—its existence should be the null hypothesis. A colleague suggests that allowing underspecified inputs in an OT grammar is problematic for various reasons. One objection is that allowing underspecification constitutes an enrichment to the model that may be as radical as, say allowing local constraint conjunction, and thus takes us out of the domain of classic OT. This objection is actually irrelevant to our purpose here, since our goal, as noted above, is not to choose which approach to chain shifts and saltations is best, but rather to present one that has not been considered before, one that uses standard-type constraints. Nonetheless, I will argue that this view that underspecification is an enrichment to a model of phonological representation is misguided: underspecification comes for free in a model that treats segments as sets of valued features.

Allowing underspecification reduces to not having in the model a stipulation that segments be complete, not insisting that segments have a value for each member of the universal feature set (the existence of which is a notion shared by any OT model with innate constraints, as well as Chomsky and Halle (1965, 1968), and also the current paper). While it is true that allowing underspecification *increases* the number of segments that a model can describe, and thus increases the number of languages (qua segment inventories) that a model can describe, it is a conceptual mistake to think of a stipulation against underspecified segments as *decreasing* complexity by providing a more restrictive model. Simple calculations show that a basic system of features, constraints, etc. can trivially yield combinatoric spaces of languages of astronomical magnitudes, even ones with more members than there are particles in the universe (around 2^{285}, what Gallistel and King (2009) call "essentially infinite"). So, the only meaningful metric for restrictiveness is in terms of the intensional characterization of a model of UG.

For example, with just twenty features, we get a system that intensionally defines 3^{20}, about 3.5 billion, segments: each segment is either lacking a specification for a given feature F, or else it is specified either +F or −F: twenty features yield twenty three-way choices. Given this set of possible segments, a language can be thought of as a segment inventory derived by answering a string of 3^{20} Yes/No questions like *Is* [u] *in the language? Is the underspecified vowel* [I] *in the language?*, and so on. The set of languages is the power set, the set of all subsets, of the universal segment inventory, and there are $2^{3^{20}}$ ways to answer that list of questions—Yes, Yes, No, Yes, No, . . . , or maybe No, Yes, No, Yes, Yes, ..., etc. This yields about 2 to the 3.5 billionth power segment inventories or languages—way more than Gallistel and King's "essentially infinite". This results from a UG with just twenty features and two values! So, a model without a stipulation that segments must be complete is more restrictive, in the scientifically relevant sense, than one without such a stipulation, yet it provides for an essentially infinite typological space (see Chomsky 2007 and Reiss 2012, 2022 for related discussion). We might even take this as phonological twist on Wilhelm von Humboldt's adage that "Language makes infinite use of finite means."[1]

A third argument for the necessity of underspecification is provided by the work of Sharon Inkelas who argues that the best analysis of sets of alternations that demand three types of lexical entries despite the existence of only two categories of surface forms is one that uses underspecification. For example, a language that has some morphemes that always surface with an [e], others that always surface with an [a], and a third set that alternate between these two vowels based on phonological conditions, is often best modeled as having /e/ in the first set, /a/ in the second set, and a vowel /A/, a non-high, non-round vowel unspecified for backness in the third set.

Inkelas provides numerous examples of patterns with this logical structure and argues against alternative analyses. She makes the following important point:

> Underspecification has been controversial since its earliest existence, drawing fire early from Stanley (1967) and more recently from Mohanan (1991); McCarthy and Taub (1992); Smolensky (1993). Aside from Stanley, however, virtually all **objections to underspecification have actually been objections to various principles designed to regulate its distribution**. These fall into [these] general categories [...]:
>
> - Markedness (universal, language-specific, or contextual); unmarked material is underspecified [refs.]
> - Redundancy; redundant feature values (as determined by the segment inventory) are underspecified [refs.]

[1]For a less dramatic, but still astronomical result, consider that, an Optimality Theoretic UG with thirty constraints in Con provides for 30! (more than 2.6×10^{32}) rankings. Nonetheless, the characterization of the model is elegantly compact and restrictive in the relevant sense.

> - Predictability: predictable material is underspecified [refs.]
>
> Critics of underspecification have found **flaws in each of these principles, a common conclusion being that underspecification is fatally tainted**. Thus Smolensky (1993) and McCarthy (1994) have claimed it to be a virtue of Optimality Theory (Prince and Smolensky 1993) that underspecification is unnecessary in the analysis of various phenomena, e.g. transparency and neutralization, once thought to require it.
>
> I argue that **underspecification is necessary, even in Optimality Theory**

The logical point bears stressing—the existence of a bad argument *for* x does not constitute an argument *against* x. For example, the claim that Saussure was a great linguist because he was Vietnamese (which he wasn't), does not support the claim that Saussure was not a great linguist.

Inkelas' extensive argumentation and analysis can be contrasted, for example, with the vague hints in other discussions in the OT literature that are biased against underspecification, not only those referred to by Inkelas, but also Tesar (2014, p.387, fn. 2): "It has been suggested that Optimality Theory reduces or eliminates the need for such underspecification (Itô et al. 1995; Smolensky 1993." There are a few problems here. First, the suggestion that an OT analysis without underspecification can be made to weakly generate the same output forms as some alternative analysis that uses underspecification cannot serve as a demonstration that OT analyses do not ever need underspecification. We would need a demonstration that, in general, OT analyses can do without underspecification, and not just cases of a certain type. Smolensky's 1993 *handout* suggests that a "wide range" of underspecification analyses *can* be recast in OT without underspecification. Another way to say that, following Inkelas' arguments, is that "some OT analyses appear to require underspecification, or, perhaps, cannot be handled by OT at all." Second, the idea of "reducing" the need for underspecification is not useful: either UG allows underspecified representations or it does not. This bias against using more than a minimum amount of underspecification will be discussed again below.

Third, the Itô et al. (1995, p. 605) paper cited by Tesar contains the following sentence: "Our analysis resolves this issue **in favor of the underspecified output structure**." This does not appear to support a ban on underspecification. More recently, in the OT tradition, Magri's (2018) "partial phonological features" analysis uses underspecification under a different label. In any case, the current paper in no way constitutes an argument that early OT should have used more or less underspecification—it explores a consequence of using it in a certain way.

It is worth pointing out that the three arguments that I have given for the acceptance of underspecification are empirically grounded, at least to the extent that scientifically proposals are generally considered to be empirically grounded. The argument from surface underspecification is based on fine acoustic analysis from various languages. The combinatoric argument is based on simple mathematical observations—it is an undeniable fact that a small model of UG can provide an essentially infinite typological space. Finally, Inkelas' examples of the necessity for a three way underlying contrast realized in a two-way surface contrast is a robust example of standard linguistic reasoning—sure, one can offer alternative means of accounting for 'exceptionality', but her representational solution is both theoretically restrictive, since it does not posit either extra features or extra mechanisms, and grounded in observation.

3 Why bother?

Why is it worth showing that beliefs about chain shifts and saltations in OT have been mistaken? After all, most phonologists probably fall into one of two camps: either they do not use OT in their work, or they use one of many available versions of OT that have developed over recent decades that are have been enriched *vis-à-vis* the "classical" OT we are concerned with here. My motivation is the belief that theory comparison should be based on sound arguments, and I intend to show here that one cannot reject classic OT on the grounds that it cannot account for these phenomena. If one wants to reject classical OT, or any other version, one should do so on legitimate grounds, but critiques have not always adhered to such standards.

For example, it is a commonplace to critique OT for the problem of needing to evaluate an infinite candidate set. However, this is probably an invalid objection if we can manage to conceive of the candidate set as partially ordered. Consider another problem of evaluating an infinite candidate set: Find the lowest prime number that is greater than 23. There are an infinite number of primes greater than 23, yet we can trivially 'filter' that set to select 29 as the prime we seek because the candidate set is ordered (because the set of integers is ordered), and we can rule out the infinite set of candidates greater than 29. This is (partially) analogous to a situation in which one OT candidate harmonically bounds (and thus eliminates) an infinite set of candidates. We need other reasons to reject standard OT if we want to do so—chain shifts and saltations are no more probative than the infinite candidate set problem.

A second reason to pursue these issues is that the solution I offer, a solution that appears to have been overlooked since the "classical period", provides a useful study of the interaction of representational and computational issues. What appears as a logical problem of constraint interaction can be circumvented by the use of underspecification. Such lessons should have wide applicability in many linguistic frameworks.

Finally, on a personal note, in the past, I not only accepted the universally held view expressed by Kager in the introduction, but I also wrote about it and exploited it in discussion of saltations (under a different name), which I claimed did not exist. I appear to have been doubly mistaken—not only do saltations appear to exist, but classical OT can generate them. So, the demonstration I provide here serves as a cathartic *mea culpa*.

4 Superficial characterization of chain-shifts and saltations

As mentioned above the chain shift and saltation problems in the OT literature are presented in terms of mappings among fully specified segments. Kirchner introduces chainshifts with this example: "If, for example, /a/ raises to [e], and /e/ raises to [i], it would appear that the /e/ → [i] raising must precede /a/ → [e] raising in the derivation; otherwise /a/ and /e/ would both neutralize to [i]." In other words, he is positing that there is a segment /a/ in some morphemes that surfaces as [a] or [e] depending on the environment, and there is a segment /e/ in some morphemes that surfaces as [e] or [i] depending on the environment. The two non-identity mappings /a/ → [e] and

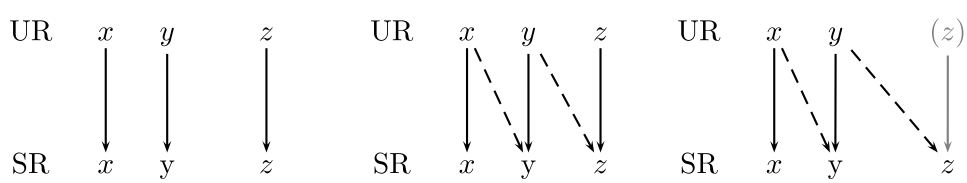
/e/ → [i] are assumed to occur in the same environment. It is typically assumed that there is also an underlying segment /i/ that consistently surfaces as [i] (unless there is a four-level chain shift), so that both of the rules are neutralizing: /a/ neutralizes with /e/ sometimes, and /e/ neutralizes with /i/ sometimes.

4.1 Segment mappings for chain shifts

So, chainshifts involve a set of identity mappings, shown by the solid vertical arrows in (1i), as well as the non-identity mappings added with the dashed arrows in (1ii):

(1) Chain shifts

i. Identity mappings   ii. All mappings   iii. Relevant mappings



In a traditional rule-based model, such a chain shift can be generated by rules (2a) and (b) applied in that order, so that underlying x does not map first to y then to z, which would occur if rule (b) fed rule (a).

(2) Counterfeeding rule ordering

Rule a. $y \rightarrow z$ / in *env*

Rule b. $x \rightarrow y$ / in *env*

Crucially x does not map to z.

In (1iii), the underlying /z/ and the /z/ to [z] mapping have been grayed out, because they are not relevant to the computational issues under discussion—rule (a) does not have to be a neutralization rule.

4.2 Segment mappings for saltations

The corresponding diagram and rule set for saltations are given in (3):

(3) Saltations

$$\begin{tikzcd}[row sep=large] \text{UR} & x \arrow[d] \arrow[drr, dashed] & y \arrow[d] & (z) \arrow[d] \\ \text{SR} & x & \mathrm{y} & z \end{tikzcd}$$

Rule $x \rightarrow z$ / in *env*

Crucially y does not map to z

Once again, it is irrelevant to our discussion whether there is an underlying $/z/$ segment that neutralizes with the output of the $/x/$ to $[z]$ mapping, so the relevant parts of the diagram are grayed out.

4.3 Segment closeness assumption

The descriptive terms ‘chain shift’ and ‘saltation’ only make sense under the assumption that segments can be partially ordered according to some phonetic and/or phonological scales. For example, the notion of saltation only makes sense if x is ‘jumping over’ y to become z in the relevant environment. This means that y must lie between the other two segments on some scale. Consistent with (sometimes only implicit) discussion in the literature, I follow Reiss (1996) in modeling such scales in terms of a notion of segment ‘closeness’ based on the subset relation of elementary set theory. Let’s use the three bilabial obstruents p,b,β to illustrate by giving their representations as sets of valued features.

(4) Illustration of closeness relation

$$\begin{array}{ccc} \text{p} & \text{b} & \beta \\ \left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ -\textsc{Voice} \\ -\textsc{Continuant} \\ \vdots \end{array}\right\} & \left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ +\textsc{Voice} \\ -\textsc{Continuant} \\ \vdots \end{array}\right\} & \left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ +\textsc{Voice} \\ +\textsc{Continuant} \\ \vdots \end{array}\right\} \end{array}$$

Every feature shared by p and β is also shared by b and β, and b and β share at least one feature that p and β do not share, namely

+Voiced, so we can say that b is closer to β than p is.[2] In simple set theoretic terms, taking each IPA symbol as the name of a set, we can express the relevant relations thus: p ∩ β ⊆ b ∩ β: the intersection of p and β is a subset of the intersection of b and β. Of course, closeness can be undefined: neither p nor β is closer to b than the other.

So, a saltation can be characterized as a mapping among segments x, y, z such that y is closer to z than x is, and x maps to z but y maps to y in the same context. Using the three labial obstruents p,b,β we can imagine a mapping from p to β that 'jumps over' b, as in (5).

(5) p b β

Discussion of chainshifts also assume a notion of closeness: "A chain shift is a situation in which sounds are promoted (or demoted) stepwise along some scale in some context" (Kager, 1999, 393). The steps on a scale are implicit in a graphic like (6).

(6) p b β

The rules of the chain shift map /p/ to the segment that is closer to it, the [b], but not to the segment that is further, [β].

There are also chain shifts that involve deletion (see McCarthy 1993 for an example), so that one end of the scale involves the absence of a segment, but we will restrict discussion here to chain shifts without deletion.

5 The problems for classic OT

5.1 Why are chain shifts a problem for Classical OT?

Chain shifts were problematic for classic OT because, as conceived, they led to ranking paradoxes. The logic is quite straightforward and we can walk through it informally using our three

[2] And also that b is closer to p than β is, since b and p share −Continuant to the exclusion of p, and they also share everything that p and β share.

labial obstruents in a hypothetical case where /p/ maps to [b] and /b/ maps to [β] in some environment.

- If /b/ → [β] is the optimal mapping, then it is more important for the output to be +CONTINUANT than to be faithful to −CONTINUANT in the environment *env*.
- This yields a constraint ranking in which the requirement to have the specification +CONTINUANT is more important than maintaining an underlying value for CONTINUANT.

  $\boxed{\text{BE+CONTINUANT} >> \text{FAITH}\{\text{CONTINUANT}\}}$

- If /p/ → [b] is the optimal mapping, then it is more important for the output to be faithful to −CONTINUANT than to be +CONTINUANT in *env*

  $\boxed{\text{FAITH}\{\text{CONTINUANT}\} >> \text{BE+CONTINUANT}}$

- So, there's a ranking paradox.

Because of this kind of ranking paradox, it is necessary to figure out how to let a winning candidate be unfaithful, but not too much. Various mechanisms, such as distance faithfulness and constraint conjunction were proposed to address this challenge.

5.2 Why are saltations a problem for classical OT?

The difficulty of modeling saltations is obvious given our discussion of closeness. A mapping from /p/ to [β] cannot be more harmonic than a mapping from /b/ to [β], since the former will have more faithfulness violations than the latter, but the two will have the exact same violations of markedness, since markedness constraints only look at potential output forms. Whereas in the case of chain shifts, the problem is to avoid being too unfaithful, with saltations, we have to avoid being just a bit unfaithful, and be extra unfaithful: a labial obstruent is unfaithful to the underlying value for CONTINUANT only if it is also unfaithful to the underlying value for VOICE.

As in the case of chain shifts, saltations lead to a ranking paradox:

(7)
- If /p/ → [β] is the optimal map, then it is more important to be +Continuant than to be faithful to −Continuant in *env*

  $\boxed{\text{Be+Continuant} >> \text{Faith}\{\text{Continuant}\}}$

- If /b/ → [b] is optimal map, then it's more important to be faithful to −Continuant than to be +Continuant in *env*

  $\boxed{\text{Faith}\{\text{Continuant}\} >> \text{Be+Continuant}}$

- So there's a ranking paradox!
- Be unfaithful ... but only if you can be *very* unfaithful! A segment should be unfaithful to Continuant if it is also unfaithful to Voice

5.3 Discussion

As mentioned above, the problem of generating both chain shifts and saltations in classic OT has been understood as a logical issue of classical constraint interaction: Kager expressed the 'fact' that "chain shifts cannot be analyzed under standard OT," and Hayes and White (2015) section 5.1 is entitled "Why saltation cannot be derived in classical Optimality Theory".[3] We are not going to discuss the various mechanisms that have been added to classical OT to get chain shifts and saltations, such as constraint conjunction, distance faithfulness, P-Maps, and others. What is important is that the 'facts' were false: classic OT can generate chain shifts and saltations without any of those mechanisms, a new result we demonstrate below.

[3] Hayes and White define 'classical' Optimality Theory as "Prince and Smolensky (1993) as modified by the Correspondence Theory of McCarthy and Prince (1995)". Other work that was already in circulation in 1995 made saltations and chain shifts possible, we shall see, but nobody seems to have realized that. I show below that one way to get these patterns is to allow underspecification and the Max and Dep faithfulness constraints for individual features proposed by Lombardi (2001/1995) and discussed by (McCarthy, 2011, 200).

6 Choice of constraints

This paper provides constructive proofs that the logic of constraint interaction in a classic OT grammar can generate surface patterns corresponding to chain shifts and saltations. The proofs are satisfied by providing a set of classic-type constraints (no local conjunction, etc.), a ranking and a lexicon that generates the relevant patterns. Note that the crucial property of chainshifts and saltations is the relationship between segments appearing in surface alternants of mophemes. For a chain shift, we need morphemes manifesting an alternation between, say [p] and [b], and others manifesting an alternation between [b] and [β]; and for a saltation, we need morphemes manifesting an alternation between, say, [p] and [β], alongside morphemes with non-alternating [b] in both relevant contexts.

Since we are interested in construction proofs in the context of classic OT, we don't need to worry about the exact set of constraints used to contruct our examples. This is consistent with the standard view in the literature that OT is fundamentally a theory of constraint interaction:

- "OT is a general framework for constraint interaction, and as such it **does not entail a particular set of constraints in** Con" (McCarthy, 2002, 17)
- "Optimality Theory is a theory of constraint interaction, not of representations. We want our deductions about OT to hold even if the theory of representations changes" (Moreton, 1999, 142).

The constraints I invoke below are consistent with many that are invoked in the literature, but I can point out that some recent OT work adopts the view—a good move, in my opinion—that putative naturalness and phonetic grounding are irrelevant for the justification of a constraint's existence:

(8) 'The pursuit of theory' (Prince 2007, 57)

> A constraint ... is a principle within a theory and, like any other principle in any other theory, is justified by its contribution to the consequences of

> that theory. Since OT is a theory of grammar, the consequences are displayed in the grammars predicted and disallowed—'typological evidence'. A constraint which cannot be justified on those grounds cannot be justified. Further, 'justifying' a constraint functionally [. . .] can have no effect whatever on its role within the theory. A constraint, viewed locally, can appear wonderfully concordant with some function, but this cannot supplant the theory's logic or compel the global outcome ('efficiency') that is imagined to follow from the constraint's presence, or even make it more likely.

In other words, constraints in a theory of phonology are justified if they are necessary to account for attested linguistic patterns. In this paper, we are not concerned with the properties of CON of the human phonological faculty, but rather with the logical properties of the classic OT model of constraint interaction.

7 The key

The crucial insight we need to model chain shifts and saltations is actually a commonplace: "The relation between a phonemic system and the phonetic record . . . is remote and complex" (Chomsky, 1964, 38). We need to reconceive of chain shifts and saltations, not as relations between surfacing segments, but in terms of derivations. The basic idea is easy to illustrate by revising the segment mapping diagrams and rules above in (1) and (3). For chain shifts, instead of mapping an underlying x to both x and y, and an underlying y to both y and z, we posit different underlying segments, ones that do not appear on the surface, as in (9), given along with rules relevant for a rule-based analysis.

(9) Revised chain shift mappings

$$
\begin{tikzcd}
\text{UR} & \alpha \arrow[d, dashed] \arrow[dr, dashed] & \beta \arrow[d, dashed] \arrow[dr, dashed] & (\gamma) \arrow[d, dashed] \\
\text{SR} & x & \mathrm{y} & z
\end{tikzcd}
$$

- $\beta \rightarrow z$ / in *env*
- $\alpha \rightarrow y$ / in *env*
- $\alpha \rightarrow x$ /elsewhere
- $\beta \rightarrow y$ / elsewhere
- Crucially $\alpha \not\rightarrow z$ / in *env*

In this revised diagram, the vertical arrows are no longer identity mappings, so I have used dashed lines for all the mappings from underlying segments to surface segments.

Let's illustrate the chain shift pattern with our three labials. Instead of deriving the alternation between [p] and [b] from /p/, I derive the two surface segments from an element that is different from both those two surface manifestations. In particular, I derive the p∼b alternation from an underlying segment /P/ which has all the features shared by /p/ and /b/, but lacks a specification for Voiced. Similarly, I derive the b∼β alternation from an underlying segment /B/ which has all the features shared by /b/ and /β/, but lacks a specification for Continuant.

Let's first list the features of all the segments under consideration for the chain shift, along with one, the Hebrew letter 'beth' which we will use later to account for saltations. Recall that we have no stipulation that segments be complete, so I've noted which features are missing from the three underspecified segments:

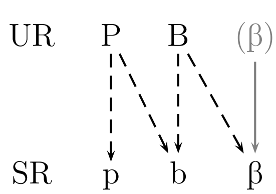
(10) Six labial obstruents

$$
\begin{array}{ccc}
\text{p} & \text{b} & \beta \\
\left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ -\textsc{Voice} \\ -\textsc{Continuant} \\ \textit{etc.} \end{array}\right\}
&
\left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ +\text{Voice} \\ -\textsc{Continuant} \\ \textit{etc.} \end{array}\right\}
&
\left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ +\textsc{Voice} \\ +\textsc{Continuant} \\ \textit{etc.} \end{array}\right\}
\end{array}
$$

$$
\begin{array}{ccc}
\text{P} & \text{B} & \text{ב (Heb. ‘beth’)} \\
\left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ -\textsc{Continuant} \\ \textit{etc.} \end{array}\right\}
&
\left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ +\textsc{Voice} \\ \textit{etc.} \end{array}\right\}
&
\left\{\begin{array}{c} +\textsc{Labial} \\ +\textsc{Anterior} \\ \textit{etc.} \end{array}\right\} \\
\boxed{\text{No }\textsc{Voice}\text{ spec.}} & \boxed{\text{No }\textsc{Cont}\text{ spec.}} & \boxed{\text{No }\textsc{Voice}\text{ or }\textsc{Cont}\text{ spec.}}
\end{array}
$$

The segment mapping diagrams and the mappings needed are given in (11):

(11) Labial chain shift mappings

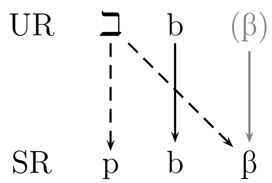


- /B/ → [β] / in *env*
- /P/ → [b] / in *env*
- /P/ → [p] /elsewhere
- /B/ → [b] / elsewhere
- Crucially /P/ $\not\rightarrow$ [β] / in *env*

It is trivial to implement these mappings with a rule system that allows for feature-filling rules but we will pursue only the OT analysis here. Note that none of the mappings (aside from the irrelevant /β/ to [β]) are identity mappings—they all involve feature-filling.

At this point, the saltation parallel to the abstract chain shift pattern in (9) should be obvious, so in (12) I just provide the concrete parallel to (11):

(12) Labial saltation mappings



- /ב/ → [β] / in *env*
- /ב/ → [p] / in *env*
- /b/ → [b] / everywhere
- Crucially /b/ ↛ [β] / in *env*

In order to get our OT derivations, we use two underspecified segments for chain shifts as indicated in (11): each segment, /P/ and /B/, is underspecified for a single feature. For saltations in (12), we use a segment /ב/ that happens to be underspecified with respect to two features.

A reader of an earlier version of this work was able to provide an alternate OT derivation for chain shifts that used a single underspecified segment. It is important to recognize two points that this person appears not to have appreciated. First, the existence of this alternative only supports the main point of this paper, that classic OT constraint interaction can generate chain shifts: the alternative *supports* my construction proof. We find the use of /P/ and /B/ to be convenient for expository reasons.

Second, once we accept the existence of underspecification in the model, it is not somehow better to make an analysis that uses less of it than one that uses more of it. Underspecification is not like aspirin—a little bit is good, but too much can be dangerous. We alluded above to this prejudice against underspecification in the discussion of Tesar's (2014) suggestion that the use underspecification should be diminished. In fact, one could argue that analyses that use more underspecification are *better* than those that use less, if one is inclined to use minimum description length as a criterion for selecting grammars among extensionally equivalent alternatives, since underspecified lexical entries contain less information than fully specified ones.

7.1 Which faithfulness constraints?

There are various types of Faithfulness constraint in the classic OT literature. For example, Ident(feature) is a simple version that treats a mapping from a segment specified as +F to one specified as −F as a violation of Faith[F], equivalent to a mapping from −F to +F. Once we accept underspecification we can adopt the Max(feature) and Dep(feature) constraints that treat, respectively, insertion and deletion of feature values as separate violations of faithfulness (see Lombardi 2001/1995 and McCarthy 2011, sec. 4.6). McCarthy's useful discussion includes this note about the use of Max vs. Ident to model feature deletion:[4]

> Both Ident(feature) and Max(feature) constraints are widely used in the literature. There is a tendency for authors to use the Ident(feature) constraints except when the Max(feature) constraints are absolutely necessary, as they seem to be in Quechua. In fact, some authors use Ident(feature) and Max(feature) constraints together in the same analysis (e.g., Lombardi 1995/2001). The attraction of Ident(feature) constraints is that they're easier to use, even if Max(feature) constraints are necessary in some cases. Eventually, this issue needs to be sorted out, since presumably phonological theory doesn't need both constraints for every distinctive feature.

A colleague has commented that an analyses of chain shifts and saltations can be done with just Ident constraints (under a particular interpretation) using my underspecification approach. However, if, as McCarthy suggests, Max constraints are sometimes strictly necessary, as in Quechua, the reanalysis is not useful—it seems reasonable to take McCarthy's discussion to mean that analyses using Ident should be seen as abbreviations of analyses using Max and Dep.

Thus, the suggestion that a solution using Ident constraints rather than Max and Dep is simpler reflects a fundamental error of reasoning. Once we acknowledge the need in *any* language, say, Quechua, for the more fine-grained Max and Dep constraints, it is a mistake to also make use of Ident just in

[4] I use the notations Max-Feature and Max(feature) interchangeably.

those languages where we can get away with doing so.[5] When pursuing simplicity in linguistics we need to consider global simplicity for Universal Grammar, not a parochial simplicity for the grammar of an individual language that happens to be under analysis. For this reason, I present analyses using Max and Dep below. Furthermore, since this paper offers a construction proof that chain shifts and saltations can be generated with classic OT constraint interaction, the existence of alternative analyses using only Ident(feature) only *strengthens* the claim.

With the Max and Dep machinery, a feature-filling mapping can violate a single Dep constraint (insertion of a single valued feature) whereas a feature-changing mapping will violate at least a Max constraint (violated by deletion) and a Dep constraint (violated by insertion).[6] I exploit this effect to generate the patterns of interest.

In addition to the Max and Dep faithfulness constraints, I adopt the following Markedness constraints for analyzing chain shifts:

(13) Markedness constraints used for chain shifts

- S(urface)-Spec: This constraint is undominated for the patterns of interest. This constraint is violated by output segments that manifest surface underspecification. In other words, all outputs are fully specified in the language. In a language with surface underspecification that perseveres from underlying underspecification (not the case here), this constraint is dominated by relevant Dep constraints. In such a case, it is worse to fill in missing values than to leave segments incomplete.
- *VTV: no stops between vowels (This is a kind of 'lenition' constraint.)
- *VSV: no voiceless obstruent between vowels (Another 'lenition' constraint.)

[5]Assuming there is no evidence in a single language that both kinds are needed for a single feature, a possibility McCarthy doesn't seem to entertain.

[6]My reasoning here is derived from the examples using Parse and Fill constraints from Inkelas (1995).

- *Fric: Obstruents should be stops; e.g., /β/ is more marked than /b/ (A typologically justified constraint.)
- Stop-Vless: Stops should be voiceless; e.g., /b/ is more marked than /p/. (Another typologically justified constraint.)

8 Chain shift derivations

Before providing tableaux for chain shifts involving the labial obstruents it is important to make two points of clarification. First, I have explained that this paper is about the logic of classic OT constraint interaction, and thus not about any particular constraints, yet I have chosen to adopt the Max and Dep constraints (instead of the simple Ident family) in light of McCarthy's discussion. For any feature F, this framework provides two constraints Max(F) and Dep(F). However, some of these constraints will not play a role in the derivations we need, so we can assume that the unlisted constraints are ranked below the ones shown in the tableaux, so low as to have no effect on the outcome of the derivation.

Second, to reiterate, this paper provides construction proofs that classic OT constraint interaction can generate chain shifts and saltations. That means that we just need to provide an input and a ranking that generates the desired pattern. We do not need to justify all the ranking decisions. In particular, I provide a single full ranking in each tableau, denoted as usual with solid lines between constraints. This does *not* imply that the given ranking is the only one that works. Put another way, a typical tableau with dashed lines between constraints represents a *set* of grammars that can generate a given pattern. It is legitimate to be interested in how much of the constraint ranking is fully determined by a given data set, and to explain how a strict ranking is justified. However, here, our interests are different—we just need to show that the pattern can be generated by at least one ranking. So, the following discussion differs from a typical OT analysis because our goals are different. We are not trying to show how much about the ranking of a particular grammar we can be sure of, but instead we are trying to show that there is a grammar that gives us what we want. To rephrase: an OT tableau with dashed lines between some unrankable constraints represents a *set* of rankings, a set of grammars. We do not use dashed lines because we adopt the classic OT assumption that a grammar is a total ranking of CON, and we are providing a single grammar that fulfills the conditions that interest us (of generating a chain shift or a saltation). Our construction proof is satisfied with a single ranking that gives the desired result.

So, the p∼b alternation will be derived from /P/, which is a bilabial stop unspecified for VOICE and the b∼β alternation derives from /B/, an underlying voiced bilabial unspecified for CONTINUANT. Recall that filling in unspecified values incurs only DEP violations: for example, a candidate with [p] corresponding to underlying /P/ incurs a violation of DEP-VOICE, but no MAX violation, because no VOICE value has been deleted. Changing a value on a feature incurs a MAX violation and a DEP violation: for example, a candidate with [β] corresponding to underlying /P/ incurs a MAX-CONTINUANT violation for deleting −CONTINUANT, as well as a violation of DEP-CONTINUANT for inserting +CONTINUANT.

The tableaux deriving outputs that surface as [pa, aba] from input /Pa, aPa/, and [ba, aβa] from underlying /Ba, aBa/, are presented below in (14-17). In all four tableaux, the two outputs with underspecified segments can never be optimal because SURFACE-SPEC is undominated. In (14), candidate (c) with [β] involves deletion of −CONTINUANT, a fatal MAX-CONT violation, as well as insertion of +CONTINUANT. Candidate (b) with [b] is eliminated by the lowest ranked constraint STOP-VLESS, since [b] is voiced.

(14) /P/ surfaces as [p] when not intervocalic

/Pa/	SURFACE-SPEC	MAX-CONT	DEP-VOICE	*VTV	DEP-CONT	*VSV	*FRIC	STOP-VLESS
☞ a. pa			*					
b. ba			*					*!
c. Pa	*!							*
d. Ba	*!							
e. ɸa		*!			*		*	

It is worthwhile pointing out that in this environment, fully specified /p/ yields the same output as /P/.[7] In fact, the mapping /p/ to [p] is more harmonic than the mapping /P/ to [p], but a /p/ to [b] mapping between vowels would lead to a ranking paradox. The underlying segment needs to yield the correct result in all environments. Under the correct ranking /apa/ maps to [apa] (see below), not to [aba] as required for the chain shift pattern. In (15), candidate (c) is eliminated by MAX-CONT because the input is −CONT and [β] is +CONT. Candidates (a) and (b) both violate constraint DEP-VOICE because they each contain a value for VOICED which is not in the input; and they both violate *VTV because they contain intervocalic stops. Candidate (a) is eliminated by *VSV, because it contains a voiceless obstruent [p] between vowels.

(15) /P/ surfaces as [b] between vowels

/aPa/	SURFACE-SPEC	MAX-CONT	DEP-VOICE	*VTV	DEP-CONT	*VSV	*FRIC	STOP-VLESS
a. apa			*	*		*!		
☞ b. aba			*	*				*
c. aβa		*!	*		*		*	
d. aPa	*!			*				*
e. aBa	*!							
f. aɸa		*!			*	*	*	

It is apparent that the optimal mapping from /apa/ given this constraint ranking would be [apa] since there would be no violation of high-ranked DEP-VOICE. So, the intervocalic environment forces us to select underlying /P/ over /p/, which was not possible in the complementary environment of (14).[8]

In (16) candidates (a,f) are eliminated by DEP-VOICE: the input is +VOICED and [p] and [ɸ] are −VOICED. Candidate (c) with [β] is eliminated by *FRIC. So the output is candidate (b) with [b].

[7] Here is the tableau with input /p/ instead of /P/ when not between vowels:

/pa/	S-SPEC	MX-CONT	DP-VOICE	*VTV	DP-CONT	*VSV	*FRIC	STOP-VLESS
☞ a. pa								
b. ba			*!					*
c. βa		*!	*		*		*	
d. Pa	*!							*
e. Ba	*!	*	*					
f. ɸa		*!			*		*	

[8] Here is the tableau with /p/ between vowels:

/apa/	SURFACE-SPEC	MAX-CONT	DEP-VOICE	*VTV	DEP-CONT	*VSV	*FRIC	STOP-VLESS
☞ a. apa				*		*!		
b. aba			*	*				*
c. aβa		*!	*		*		*	
d. aPa	*!			*				*
e. aBa	*!							
f. aɸa		*!			*	*	*	

(16) /B/ surfaces as [b] when not intervocalic

/Ba/	Surface-Spec	Max-Cont	Dep-Voice	*VTV	Dep-Cont	*VSV	*Fric	Stop-Vless
a. pa			*!		*			
☞ b. ba					*			*
c. βa					*		*!	
d. Pa	*!							*
e. Ba	*!							
f. ɸa			*!		*		*	

In (17) candidates (a,f) are eliminated by Dep-Voice: the input is +Voiced and [p] and [ɸ] are −Voiced. Candidate (b) with [b] is eliminated by *VTV. So the optimal candidate is (c) with [β].

(17) /B/ surfaces as [β] between vowels

/aBa/	Surface-Spec	Max-Cont	Dep-Voice	*VTV	Dep-Cont	*VSV	*Fric	Stop-Vless
a. apa			*!	*	*	*		*
b. aba				*!	*			*
☞ c. aβa					*		*	
d. aPa	*!			*				*
e. aBa	*!							
f. aɸa		*!	*		*	*	*	

We have now generated a chain shift pattern in which one underlying segment surfaces as [b] intervocalically and [p] elsewhere, and another underlying segment surfaces as [β] intervocalically and [b] elsewhere. We have done this without conjoined contraints or distance faithfulness, using only what Hayes and White call "classical OT" interactions. Kager's universally accepted statement that "chain shifts cannot be analyzed under standard OT" cannot be maintained.

9 Saltation derivations

In order to simplify the tableaux for saltations, I eliminate all forms with surface undespecification and assume that Surface-Spec is undominated, as for the previous language. We need one additional markedness constraint:

(18) Stop-Ons: Onsets should contain stops. An onset without a stop incurs a violation.

The relevant faithfulness constraints are drawn from the same Max and Dep families used for the chain shift analysis.

We begin by giving a ranking that will perform the identity mapping /b/ → [b] both intervocalically and elsewhere.

(19) b → b / V __V

/aba/	Max-Voice	Max-Cont	*VTV	*VSV	Dep-Voice	Dep-Cont	Stop-Ons	Stop-Vless
☞ a. aba			*					*
b. apa	*!		*	*	*			
c. aβa		*!				*	*	
d. aɸa	*!	*		*	*	*	*	

(20) b →b when not intervocalic

/ba/	Max-Voice	Max-Cont	*VTV	*VSV	Dep-Voice	Dep-Cont	Stop-Ons	Stop-Vless
☞ a. ba								*
b. pa	*!				*			
c. βa		*!				*	*	
d. ɸa	*!	*			*	*	*	

The generation of non-alternating [b] is straightforward, so we need not walk through it. Now let's turn to the [p]∼[β] alternation.

In (21) and (22) the input contains ℶ, and now Max constraints are irrelevant, because there are no values to delete for Voice and Continuant. On the other hand, every candidate violates the Dep constraints—each inserted valued feature incurs a violation (and we have left out the forms with surface underspecification that would be eliminated by Surface-Spec). The constraints that favor stops in onsets (over fricatives) and that favor voiceless stops over voiced ones in general, ensure that the winning candidate has either [β] or [p].

When ℶ occurs between vowels, as in (21), the candidates with [b] and [p] are excluded by a violation of *VTV, and the candidate with the voiceless fricative is excluded by *VSV. The 'leap' to [β] can surface, despite the insertion of two valued features, +Voiced and +Continuant.

(21) ℶ →β / V __V

/aℶa/	Max-Voice	Max-Cont	*VTV	*VSV	Dep-Voice	Dep-Cont	Stop-Ons	Stop-Vless
a. aba			*!		*	*		*
b. apa			*!	*	*	*		
☞ c. aβa					*	*	*	
d. aɸa				*!	*	*	*	

When ℶ does not occur between vowels, as in (22), the markedness constraints that make reference to intervocalic position are irrelevant. Of course, the Max constraints are still irrelevant, since no valued feature is deleted. The candidate with [p] is optimal, since [b] violates the constraint demanding that stops be voiceless.

(22) ב →p / #___

/בa/	Max-Voice	Max-Cont	*VTV	*VSV	Dep-Voice	Dep-Cont	Stop-Ons	Stop-Vless
a. ba					*	*		*!
☞ b. pa					*	*		
c. βa					*	*	*!	
d. ɸa					*	*	*!	

White's (2017) claim that "saltations cannot be derived in 'classical OT' " cannot be maintained.

10 Irrelevance of the Richness of the Base

Several readers have objected that the preceding analyses are incomplete without discussion of the Richness of the Base (RotB) (e.g., Smolensky, 1996; Kager, 1999)). As stated above, this paper is not the place to delve into a full discussion of the RotB as a foundational assumption of Optimality Theory. I do offer the following observations, nonetheless.

First, RotB is supposed to reflect the idea that the surface inventory of a language is fully determined by the constraint ranking. To illustrate in simple terms, if we could somehow feed a morpheme containing a click into an English-type grammar, the constraint ranking would guarantee that the click would never surface, and the optimal candidate would contain, say, [k]. As suggested above, this issue is orthogonal to the matter of chain shifts and saltations, where the concern is with *pairs* of segments surfacing in related forms—say, [p] and [b] surfacing in alternants of the same morpheme depending on phonological context, while [b] and [m] surface in the alternants of other morphemes. The question of how yet other input segments might surface is irrelevant to the explorations here about alternations.

Second, RotB is predicated on the false claim that loanwords do not introduce new segments of sequences into a language. French was the source of the [ʒ] in *pleasure* and *treasure*—only the corresponding voiceless sound [ʃ] occurred in English beforehand.

In recent years Yiddish loanwords to English have introduced the [ʃm] and [ʃl] sequences of *shmooze* and *shlep*. Some speakers even produce formerly "banned" [sr] in words like *Sri Lanka* and *Sriracha sauce*. So, there is no reason to consider RotB to be valid.

Finally, RotB, according to Smolensky (1996), has a logical consequence an initial ranking of the OT hierarchy with all markedness constraints M, ranked above all faithfulness constraints, F. This initial ranking has been demonstrated to be untenable in the pages of this journal (Hale and Reiss, 1998), and thus RotB can be rejected on simple logical grounds: RotB implies M >> F (Smolensky's demonstration); M >> F is false as a characterization of the initial state (Hale and Reiss); therefore, RotB cannot be maintained (by *modus tollens*; see AUTHOR (submitted) for details).

In light of these arguments that RotB is untenable on the one hand, and irrelevant to the analysis of chain shifts and saltations—even if it were tenable, on the other, we are justified in discussing RotB no further with regard to the topic of this paper.

11 Conclusion

The preceding discussion provides construction proofs that classic OT grammars, without local conjunction or other machinery beyond that introduced by 1995, can generate both saltations and the chain shift patterns associated with counterfeeding rule ordering. Dispelling the myth that classic OT grammars cannot, by their nature, generate such patterns allows theory comparison to advance on firmer ground. We now know that the capacity to generate chain shifts and saltations does not distinguish between classical OT and derivational approaches.

The apparent computational problems have a simple representational solution. The potential objection that I have just chosen the input segments in order to make things 'work' is not valid. Given the parameters of the model, of course we want to posit inputs that yield the right results. A better way to think of the issue is the following: If we assume a model that allows binary features and underspecification, as well as MAX-F and DEP-F constraints, is there a lexicon and a constraint ranking that a learner can posit that will generate the observed pattern? It appears that the lexical entries and classic OT grammars presented here provide an affirmative answer.[9]

Since we have defined chain shifts and saltations in terms of logical structure, the toy examples used here can be applied to natural language data that is consistent with the structure posited. For example, the ɛ-e-i chain shifts found in several southern Italian metaphony systems (Krämer, 2016) can be analyzed in a manner parallel to the p-b-m examples, since the same closeness relations hold among the segments. Examples of saltations are less common, but "in Campidanian Sardinian, intervocalic [p] is realized as [β]—leaping over [b], which does not alternate" (Hayes and White, 2015). The key to the reconceptualization I have provided is the use of underspecification in underlying segments. While many scholars have made use of underspecified representations for various purposes, it was primarily the rigorous and insightful work of Sharon Inkelas (Inkelas and Cho, 1993; Inkelas, 1995) that made possible the simple new results offered here, results that confirm that "[t]he relation between a phonemic system and the phonetic record . . . is remote and complex".

References

Choi, John D. 1992. Phonetic underspecification and target interpolation: An acoustic study of Marshallese vowel allophony. Doctoral Dissertation, UCLA, Los Angeles.

Chomsky, Noam. 1964. Formal discussion in response to W. Miller and S. Ervin. In *The Acquisition of Language*, ed. Ursula Bellugi and Roger Brown, 35–39. Chicago: University of Chicago Press.

[9]This result has led colleagues to offer other solutions (p.c.). The existence of these alternatives only strengthens my construction proof demonstrations. The evaluation of competing approaches to generating chain shifts and saltations under classic OT assumptions is beyond the scope of this paper.

Chomsky, Noam. 2007. Approaching UG from below. In *Interfaces + recursion = language?: Chomsky's minimalism and the view from syntax-semantics*, ed. Uli Sauerland and Hans-Martin Gärtner, 1–24. Berlin: Mouton de Gruyter.

Chomsky, Noam, and Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of Linguistics* 1:97–138.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Gallistel, C. Randy, and Adam Philip King. 2009. *Memory and the computational brain: Why cognitive science will transform neuroscience*. Chichester, West Sussex, UK: Wiley-Blackwell.

Gnanadesikan, Amalia Elisabeth. 1997. *Phonology with ternary scales*. University of Massachusetts Amherst.

Hale, Mark, and Charles Reiss. 1998. Formal and empirical arguments concerning language acquisition. *Linguistic Inquiry* 29:656–683.

Hayes, Bruce, and James White. 2015. Saltation and the p-map. *Phonology* 32:267–302.

Inkelas, Sharon. 1995. The consequences of optimization for underspecification. In *Proceedings of the North East Linguistic Society 25*, ed. Jill Beckman, 287–302. University of Pennsylvania: Graduate Linguistic Student Association.

Inkelas, Sharon, and Young-mee Yu Cho. 1993. Inalterability as prespecification. *Language* 69:529–574.

Itô, Junko, Armin Mester, and Jaye Padgett. 1995. Licensing and underspecification in optimality theory. *Linguistic Inquiry* 26:571–614.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Keating, Patricia. 1988. Underspecification in phonetics. *Phonology* 5:275–292.

Kirchner, Robert. 1996. Synchronic chain shifts in optimality theory. *Linguistic Inquiry* 27:314–350.

Krämer, Martin. 2016. *Metaphonic chain shifts, vowel height, and markedness*, volume 20, 277. Walter de Gruyter GmbH & Co KG.

Lombardi, Linda. 2001/1995. Why place and voice are different: Constraint-specific alternations in optimality theory. *Segmental phonology in Optimality Theory: Constraints and representations* 13–45.

Magri, Giorgio. 2018. Output-drivenness and partial phonological features. *Linguistic Inquiry* 49:577–598.

McCarthy, John. 1994. On coronal "transparency". Talk at Trilateral Phonology Weekend II. Stanford University.

McCarthy, John. 2002. *A thematic guide to optimality theory*. Cambridge: Cambridge University Press.

McCarthy, John, and Alan Prince. 1995. Faithfulness and reduplicative identity. In *University of massachusetts occasional papers in linguistics, vol. 18*, ed. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbancyk, 249–384. Amhert, MA: GLSA, University of Massachusetts at Amherst.

McCarthy, John, and Alison Taub. 1992. Review of paradis and prunet 1991. *Phonology* 9:363–70.

McCarthy, John J. 1993. The parallel advantage: Containment, consistency, and alignment. Paper presented at Rutgers Optimality Workshop-1, Rutgers University, October 1993.

McCarthy, John J. 2011. *Doing optimality theory: Applying theory to data*. John Wiley & Sons.

Mohanan, K.P. 1991. On the bases of Radical Underspecification. *Natural Language and Linguistic Theory* 9:285–325.

Moreton, Elliott. 1999. Non-computable functions in optimality theory. Ms., University of Massachusetts, Amherst. Rutgers Optimality Archive ROA-364.

Prince, Alan. 2007. The pursuit of theory. In *The Cambridge Handbook of Phonology*, ed. P.V. De Lacy. Cambridge University Press.

Prince, Alan, and Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar (technical report)*. New Brunswick, NJ: Rutgers Center for Cognitive Science.

Reiss, Charles. 1996. Deriving an implicational universal in a constrained OT grammar. In *Proceedings of North East Linguistic Society*, ed. Kiyomi Kusumoto, 303–318. Harvard University and MIT: Graduate Linguistic Student Association.

Reiss, Charles. 2012. Towards a bottom-up approach to phonological typology. In *Towards a biolinguistic understanding of grammar*, ed. A.M. di Sciullo, 169–191. Philadelphia: John Benjamins.

Reiss, Charles. 2022. Plastics. In *Ha! Linguistic Studies in Honor of Mark R. Hale*, ed. Laura Grestenberger, Charles Reiss, Hannes A. Fellner, and Gabriel Z. Pantillon, 327–330. Vienna: Dr. Ludwig Reichert Verlag.

Smolensky, Paul. 1993. Harmony, markedness, and phonological activity. Handout of keynote address, rutgers optimality workshop 1. ROA-87.

Smolensky, Paul. 1995. On the internal structure of the constraint component Con of UG Handout of talk at UCLA. Handout of talk at UCLA. Online at ROA-86., April 1995.

Smolensky, Paul. 1996. The initial state and 'richness of the base'in optimality theory. *Rutgers Optimality Archive* 293.

Stanley, R. 1967. Redundancy rules in phonology. *Language* 393–436.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. *The role of speech perception in phonology* 219–250.

Steriade, Donca. 2008. The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In *The nature of the word: Studies in honor of Paul Kiparsky*, ed. K. Hanson and Sharon Inkelas, 151–180. MIT Press.

Tesar, Bruce. 2014. *Output-driven phonology: Theory and learning*. Cambridge, U.K.: Cambridge University Press.

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

Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: corpus and survey evidence from tagalog infixation. *Language* 277–316.