Harmony in Harmonic Grammar by Reevaluating Faithfulness[*]

Charlie O'Hara

University of Southern California

1.   Introduction

In recent years, a rash of research has begun to show some analytical benefits that the weighted constraints of Harmonic Grammar (HG: Legendre et al. 1990, 2006, Pater 2009b, 2016) offer over the ranked constraints of Optimality Theory (OT: Prince & Smolensky 1993/2004, McCarthy & Prince 1995). The major difference in constraint interaction between the theories comes from gang up effects. While gang effects often allow complex data to be analyzed elegantly with fewer, more basic constraints than needed in OT, (as in Pater (2009a), Potts et al. (2010), Jesney (2011, 2016) among many others) some observed patterns are difficult to analyze in HG using mainstream representational assumptions. Here in this paper, we examine unbounded harmony processes, which are particularly difficult to model in parallel Harmonic Grammar.

The critical difficulty for modeling harmony in HG is that in harmony patterns one trigger segment can cause a featural change in a potentially unbounded number of target segments. This is particularly dangerous in *control-harmony* systems (a la Baković (2000)), where one privileged position's underlying feature value spreads, regardless of whether it is [+F] or [−F]

Consider Tuvan, a Turkic language with backness harmony (Harrison 2000, Rose & Walker 2011). If the first vowel of the stem is [+back], so is the rest of the word, including the suffixes, as in (1); but if it is [−back], all the other vowels must be [−back] as well, (2).

(1)   a.   at-tar-ɯm-dan   'name'PL-1-ABL
        b.   udu-ba-dɯ-m   'sleep'NEG-PST.II-1

(2)   a. is-ter-im-den   'footprint'PL-1-ABL
          b. esker-be-di-m   'notice'NEG-PST.II-1

By the principle of Richness of the Base (Prince & Smolensky 1993/2004), all linguistically possible inputs must map to a word that could be grammatical in Tuvan. Since all native Tuvan words maintain the same backness throughout all of their vowels, this cannot be explained through prespecification without some constraints on the input. With a high-ranked harmony driving markedness constraint[1]

    With just simple faithfulness constraints, OT predicts *majority rules* pathologies as in (3). The backness of the word is entirely dependent on whichever feature there were more of in the input. Majority Rules languages have long been considered pathological (Lombardi 1999, Baković 2000, Riggle 2004, Heinz & Lai 2013), and experimental artificial language learning results from Finley (2008) back up this claim. This can be solved by adding a positional faithfulness constraint to select a specific vowel (often, and for Tuvan, the initial or the stem) that is always ranked above the general constraint, keeping the privileged vowel faithful to serve as the trigger (4) (Beckman 1998).[2]

(3)   *Majority Rules in OT*

/+ - - -/	AGREE(F)	ID(F)
a. + + + +		∗∗∗W
☞ b. - - - -		∗
/+ - + +/	AGREE(F)	ID(F)
☞ c. + + + +		∗
d. - - - -		∗∗∗W

(4)   *Positional Faithfulness Can Solve Majority Rules in OT*

/+ - - -/	AGREE(F)	ID(F)/σ₁	ID(F)
☞ a. + + + +			∗∗∗
b. - - - -		∗W	∗L
/+ - + +/	AGREE(F)	ID(F)/σ₁	ID(F)
☞ c. + + + +			∗
d. - - - -		∗W	∗∗∗W

This problem is not so easily resolved in HG. HG is by its nature a more democratic system than OT. While a violation of a high weighted constraint has more power than a violation of a lower weighted constraint, this difference is finite. In other words, a violation in the privileged position basically gets some finite number more votes than any violation else-

---

[1]AGREE(F) is used here, but any harmony driver can achieve this.

[2]This solution is not without its issues, as it requires a restriction of the factorial typology; and in cases where no segment is most privileged (i.e. voicing assimilation in coda clusters), can still create majority rules effects (Baković 2000), as well as having issues handling systems with segments that block or are transparent to harmony.

where, but once enough violations of the lower weighted constraints join the gang, they can overpower the high weighted constraint.

In (5), we can see that the initial segment's faithulness violation is worth $w(\text{ID}(F)/\sigma_1) + w(\text{ID}(F))$, or 4; where the other segments' violations are worth just $w(\text{ID}(F))=1$. Thus the initial syllable controls the direction of harmony when disagreeing segments outnumber agreeing segments by less than four (as they do with /+ − −−/). A majority rules type harmony occurs when there are more than 4 disagreeing segments, as with /+ − − − −−/.

(5)   *HG is more democratic than OT*

/+ − − −/	AGREE(F)	ID(F)/σ_1	ID(F)	H
	$w = 5$	$w = 3$	$w = 1$	H
☞ a. + + ++			-3	-3
b. − − −−		-1	-1	-4
/+ − − − − −/	AGREE(F)	ID(F)/σ_1	ID(F)	H
c. + + + + ++			-5	-5
☞ d. − − − − −−		-1	-1	-4

These number sensitive effects are quite problematic for theories of weighted constraints. They are caused by unbounded tradeoffs (Legendre et al. 2006, Pater 2016), which will be discussed in section 2. This paper will show that this unbounded tradeoff cannot be solved solely through reformulation of markedness constraints (section 3), and proposes a strategy that modifies the faithfulness constraints and the representational assumptions (section 4). Section 5 closes with discussion of implications and issues.

2.   Unbounded Tradeoff Problem

As mentioned above, the critical locus of difference between OT and HG is in gang effects. Pater (2016) frames these in terms of asymmetric tradeoffs. Tradeoffs are apparent when comparing two output candidates.

(6)   A tableau features an *asymmetric trade-off* iff there are two candidates a, b where the violation difference between a and b is such that n violations of one constraint $C1$ favor a, and some $m > n$ violations of another constraint(s) $C2$ favor b.

On their own, on a one tableau scale, asymmetric tradeoffs are relatively uninteresting. If every candidate pair across the language has the same asymmetric tradeoff on $C1$ and $C2$, then we only care whether $nw(C1) > mw(C2)$, or vice versa. We can call this a $n − m$ *tradeoff*. However, if we see other tradeoffs for the same two constraints, say a $n − p$ tradeoff ($p \neq m$), there are more predicted languages in HG than OT given the same set of constraints. Consider ID(F) and ID(F)/σ_1 in (5). There is both a 3-1 tradeoff for the first input, and a 5-1 tradeoff for the second. In OT, we have a two types of languages, $C1 \gg C2$ and $C2 \gg C1$. However, in HG two weighting conditions now matter, that between $nw(C1)$ and $mw(C2)$, and that between $nw(C1)$ and $pw(C2)$.

Typically these asymmetric tradeoffs have a restricted effect, often capturing things that OT misses, say where one new violation of a faithfulness constraint can be traded for a gang up of violations of multiple markedness constraints, in some contexts, but just one violation in others. However with most sets of constraints, there are a *bounded* number of crucial weighting conditions, since one violation of a constraint can only trade for some bounded number of violations of other constraints.

Yet, with the appropriate set of constraints, unbounded tradeoffs can be produced by HG, where an unbounded number of $1 - n$ tradeoffs occur, between two constraints. Legendre et al. (2006) discuss an unbounded tradeoff between MAINSTRESSRIGHT, which is a gradient ALIGN constraint that is violated by each syllable that intervenes between the main stressed syllable and the right edge of the word; and WEIGHT-TO-STRESS which assigns a violation mark to a stressed syllable if it is not heavy.

(7)   *Legendre et al. (2006)'s Unbounded Tradeoff*

ban$\sigma_n$ta	MAINSTRESSRIGHT	WEIGHT-TO-STRESS	HARMONY
	$w = A$	$w = B$	
a. 'ban.σ_n.ta	-1-n		-(1+n)(A)
b. ban.σ_n.'ta		-1	-B

In (7), σ_n represents n light syllables. We can see here that stress must either fall on the leftmost syllable, satisfying WEIGHT-TO-STRESS, or on the right most syllable, satisfying MAINSTRESSRIGHT. Note that MAINSTRESSRIGHT incurs more violations dependent on how long the word is, creating asymmetric tradeoffs, where some larger number of violations of MAINSTRESSRIGHT are traded for one violation of WEIGHT-TO-STRESS. In OT, this is no problem, as the winning candidate is dependent only on constraint ranking, not on word length. However in HG, the gang effects weaken WEIGHT-TO-STRESS's attempts at dominance. No matter how high weighted WEIGHT-TO-STRESS is and how low weighted MAINSTRESSRIGHT is, there exists some number n so that $(1 + n)A > B$. In other words, since constraint weights are positive numbers, there is no weighting so that all $1 - n$ tradeoffs prefer the same candidate.

This creates pathological counting languages, where words of this shape that have less than n syllables have quantity sensitive stress, but longer words have rightmost stress. Since there is no theoretical upper bound on word size, this creates an infinite typology in HG, for all integers $n \geq 0$, since there are an infinite number of distinct weighting conditions created by all the $1 - n$ tradeoffs.

Pater (2016) notes that this example, with these sets of constraints, is not particularly persuasive because other stress constraints could be considered, and these constraints are controversial even in OT. However, in stem control harmony, we come across the exact same issue. Assume a high weighted harmony driver, like AGREE(F), outweighs the faithfulness constraints, restricting the potential optimal candidates to the fully harmonic ones. In (8), we see an asymmetric tradeoff between the positional and general faithfulness constraints. This again is an unbounded tradeoff, because as the word length increases unboundedly, so does the number of more violations of the general faithfulness constraint.

These constraints predict an infinite typology of counting languages, where majority rules effects kick in once there are more than n segments which disagree with the privileged segment, but control harmony appears elsewhere.

(8)   *Unbounded Tradeoff in Harmony*

$+-_n-$	AGREE(F)	ID(F)/σ_1	ID(F)	HARMONY
	$w = A$	$w = B$	$w = C$	
a. $++_n+$			-1-n	-(1+n)(C)
b. $--_n-$		-1	-1	-(B+C)

There are two options to solve this issue in HG: find a harmony driving markedness constraint that obviates this issue and handles it in a phenomenon specific way (which we will see in section 3 to be impossible), or reevaluate the faithfulness constraints and the representations that underlie them, leading to much more global theoretical modifications.

3.   Majority Rules under any Markedness Constraint

The route of least resistance would be to modify the markedness constraint. Throughout most work on OT and its relatives, the majority of the research energy has been used discovering more economic or more grounded incarnations of markedness constraints. This has some obvious causes: markedness constraints are typically phenomenon specific, whereas the same faithfulness constraint would be used for any phenomenon where it could potentially serve as a possible repair. Thus, a modification of faithfulness theory likely has wide reaching effects across all phonological phenomena, and may in turn require modifications to other markedness constraints elsewhere.

However, no modification of markedness constraints can fix the majority rules problem in HG. This is truly an issue with IDENT(F) constraints in a weighted constraint framework.

To prove this, we will rely on the notion that a HG (or OT) system can only make use of markedness or faithfulness constraints: Markedness constraints are defined as incurring the same violations for the given output, regardless of the input; faithfulness constraints are not violated by an input that maps to itself. This assumption is necessary to prevent circular, or infinite chain shifts (Moreton 1999).

In order to prevent length based effects, we must show that the grammar is such that for any $n > m > 0$, if an initial segment spreads to m segments, it must spread to n segments as well. Since the weights of ID(F)/σ_1 and ID(F) are some positive real numbers, A and B respectively; there exists some integer k, where $k+1$ is the lowest integer so that $k+1$ violations of ID(F) outweigh one violation of ID(F)/σ_1, (i.e $(k+1)B \geq A \geq kB$). Thus, for any word with more than $k+1$ syllables, the majority rules candidate is preferred by the faithfulness constraints. The tableau in (9) shows this for any word with $i+k+1$ syllables, where i is any positive integer.

(9)   *Harmonic difference over faithfulness*

$+-_{i+k+1}$	$\text{ID(F)}/\sigma_1$	ID(F)	H
	$w = A$	$w = B$	
a. $++_{i+k+1}$		-(i+k+1)	$-B*(i+k+1)$
☞ b. $--_{i+k+1}$	-1	-1	$-(B+A)$

Thus the harmonic difference between $/+-_{i+k+1}/\rightarrow[++_{i+k+1}]$ and $/+-_{i+k+1}/\rightarrow[--_{i+k+1}]$ over the faithfulness constraints is $-B*(i+k+1)-(B+A)$. Since $B(k+1) \geq A$, this difference must be greater than or equal to $i(B)$.

(10) $B*(i+k+1) - (B+A) \geq (B(i+1)+A) - (B+A) = i(B)$

Since faithfulness prefers the majority rules candidate by $i(B)$, if markedness changes could avoid majority rules pathologies, whatever markedness solution we use must disfavor the majority rules candidate by more than $i(B)$.

In tableau (11) the MARKEFFECT represents the collapsed effect of all the markedness constraints in the system. Any harmony-driving constraint, as well as any other markedness constraint that we attempt to use to avoid majority rules pathologies are included here. In the violation boxes I write the relative harmony scores caused only by markedness constraints to each candidate. Since the effect of markedness surpasses the harmonic difference over faithfulness of the two candidates, all inputs of any length harmonize to an initial syllable with a [+F] feature.

(11)   *Markedness must prefer* $++_{i+k+1}$

$+-_{i+k+1}$	MARKEFFECT	$\text{ID(F)}/\sigma_1$	ID(F)	H
		$w = A$	$w = B$	
☞ a. $++_{i+k+1}$	0		-(i+k+1)	$-B*(i+k+1)$
b. $--_{i+k+1}$	$< -i(B)$	-1	-1	$< -(B(i+1)+A)$

However, this now fails us for [−F] harmony. Since markedness constraints make no reference to the input, the markedness effect must be the same for $/+-_{i+k+1}/\rightarrow[--_n]$ and $/-+_{i+k+1}/\rightarrow[--_n]$ as seen in (12). This means that the markedness effect can join forces with the gang to defeat the privilege of the initial syllable.

(12)   *Markedness fails to get* $/-+_{i+k+1}/ \rightarrow [--_{i+k+1}]$

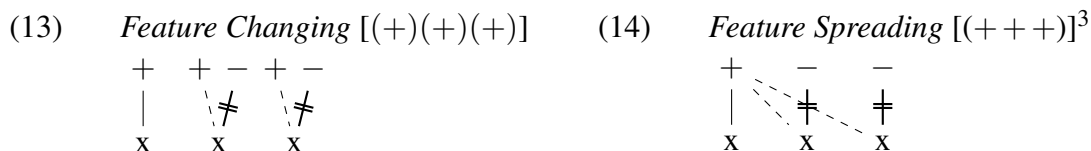| $-+_{i+k+1}$	MARKEFFECT	$\text{ID(F)}/\sigma_1$	ID(F)	H
		$w = A$	$w = B$	
☞ a. $++_{i+k+1}$		-1	-1	$-(A+B)$
b. $--_{i+k+1}$	$-i(B)$		-(i+k+1)	$-(B(2i+k+1))$

MARKEFFECT represents the overall effect of markedness constraints abstractly, regardless of which actual constraints we use. Thus, whether we use AGREE or ALIGN or any other type of harmony driving markedness constraint (even those that have not been considered),

there is no way to get stem-control harmony without getting majority rules effects. If this problem is going to be resolved it must be through faithfulness constraints.

4.   Modifying Faithfulness and Representations

The unbounded tradeoff that leads to the length-sensitive majority rules-like effect is intrinsically linked to the faithfulness constraints. If no markedness constraint can solve this problem, the only way to solve it in HG is to modify how faithfulness violations are incurred.

The critical issue is that the IDENT(F) family of constraints model feature changing, which is not necessarily the same thing as feature spreading. By adding a level of abstract representational structure, we can compare feature changing, where each underlying feature must be replaced by a new feature (13); with feature spreading where one feature just needs to link to other segments who have lost their underlying feature (14).

(13)   *Feature Changing* $[(+)(+)(+)]$ (14)   *Feature Spreading* $[(+++)]^3$

$$
\begin{array}{cccc}
+ & + - & + - \\
| & \diagdown\!\!\!\!\!+ & \diagdown\!\!\!\!\!+ \\
\text{x} & \text{x} & \text{x}
\end{array}
\qquad\qquad
\begin{array}{cccc}
+ & - & - \\
| & + & + \\
\text{x} & \text{x} & \text{x}
\end{array}
$$

As a simple replacement for IDENT(F), MAX/DEP($\pm$F) can be used instead, maintaining these distinctions.

(15)   MAX($\pm$F)/DEP($\pm$F): Assign a violation mark for any $[\pm$F] feature in the input(/output) that does not have an output(/input) correspondent.

Constraints similar to these have been previously used for privative featural accounts, first suggested in (McCarthy & Prince 1995, ff. 49), and considered in greater detail in Lombardi (2001) [4]. My MAX/DEP($\pm$F) account is crucially different as it uses binary features.

With privative features, MAX/DEP(F) obtain an asymmetric effect, similar to that of the IDENT-IO(+F) and IDENT-IO(-F) of Pater (1999). In (16), the [+F] value of F is treated as the privative feature value while using the MAX/DEP(F) constraints, showing that the privative constraints do not assign the same violations for deletion of a $[-$F] feature as deletion of a $[+$F] feature, whereas the binary featured MAX($\pm$F) does. This difference becomes critical when we see how MAX/DEP($\pm$F) avoids the unbounded tradeoff.

---

[3]Throughout the rest of this paper, parentheses () will be used to represent a span of segments linked to the same feature.

[4]Also see Causley (1997), Myers (1997), Walker (1997).

(16)   *Difference between privative (+F=F, −F=∅) and binary* MAX *feature constraints*

(+)(−)	MAX(F)	MAX(±F)
a. (+)(+)		*
b. (++)		*
(−)(+)	MAX(F)	MAX(±F)
c. (−)(−)	*	*
d. (−−)	*	*

MAX/DEP(±F) maintains the difference between feature changing and feature spreading. Unlike IDENT, DEP(±F) is violated more by a segment that is linked to an epenthetic feature rather than linked to a spreading feature, (17).

(17)   *Difference between* IDENT *and* DEP(±F)

(+)(−)	IDENT(F)	DEP(±F)
a. (+)(+)	*	*
b. (++)	*	

On the other hand, outputs with a spread feature, even if the target segments underlyingly had the same specification for the feature violate MAX(±F) but not IDENT(F), (18). This is because a segment does not violate IDENT(F) if it is linked to a different feature, but its specification is the same as in the input.

(18)   *Difference between* IDENT *and* MAX(±F)

(+)(+)	IDENT(F)	MAX(±F)
a. (+)(+)		
b. (++)		*

    In order to get the most out of the difference between IDENT(±F) and MAX(±F), it is crucial that the harmony driving markedness constraint always trades one-to-one with MAX(±F). This is achieved by using a markedness constraint that prefers adjacent segments linked to the same feature over adjacent segments each linked to separate features, even if the specifications are the same. Following McCarthy (2011) and Mullin (2011), who use a similar constraint (though associated with privative features), I will call this constraint SHARE(±F), (19).

(19)   SHARE(±F): Assign a violation mark for any segment that is adjacent to another segment that is linked to a different [±F] feature.[5]

Now, if all underlying features are linked to just one vowel, both fully harmonic candidates ((20)-a,b) must have violated MAX(±F) once for each feature in the word (except for one,

---

[5]This definition should be elaborated depending on theoretical assumptions. Perhaps we mean adjacent *on some tier*; and possibly this constraint would need to be relativized to certain domains, like the prosodic word or the stem or the phonological phrase.

the spreader). The only thing that distinguishes the candidates is whether the positional MAX($\pm$F)/σ_1 is violated, giving us initial control harmony. Thus, the correct winner (20)-a, harmonically bounds the majority rules candidate (20)-b; thus in categorical HG there is no weighting of constraints where we can get majority rules.

(20)   *No majority rules with* MAX/DEP($\pm$F).

/+−$_n$/	SHARE($\pm$F)	MAX($\pm$F)	MAX ($\pm$F)/σ_1	H
	$w = 4$	$w = 3$	$w = 1$	
☞ a. $(++_n)$		-n		-3n
b. $(--_n)$		-n	-1	-3n-1
c. $(+)(--_{(n-1)})$	-1	-(n-1)		-3n-1
d. $(+)(-)_n$	-n			-4n

The partially harmonizing candidate (20)-c is collectively harmonically bounded by the control harmonizing candidate and the fully faithful candidate (20)-d, as evidenced by the fact that the winner is chosen solely by which constraint is weighted higher, SHARE($\pm$F) (20) or MAX($\pm$F) (21). This result is independent of word length, as the tradeoff between the two constraints is symmetric (both have *n* violations).

(21)   *No harmony with high weighted* MAX($\pm$F)

/+−$_n$/	MAX($\pm$F)	SHARE($\pm$F)	MAX ($\pm$F)/σ_1	H
	$w = 4$	$w = 3$	$w = 1$	
a. $(++_n)$	-n			-4n
b. $(--_n)$	-n		-1	-4n-1
c. $(+)(--_{(n-1)})$	-(n-1)	-1		-4n+1
☞ d. $(+)(-)_n$		-n		-3n

This solution works equally well to derive harmony to [−F] (22); unlike in a system with privative features and MAX/DEP(F), none of the constraints we are using differentiate between feature values when incurring violation marks. In (23), we can see that using privative [F], the initial control candidate (23)-a is harmonically bounded by the candidate (23)-b. With MAX/DEP, privative features can only predict dominant-recessive harmony.

(22)   DEP/MAX($\pm$F) *can get spread of* −F

/−+$_n$/	SHARE($\pm$F)	MAX($\pm$F)	MAX ($\pm$F)/σ_1	H
	$w = 4$	$w = 3$	$w = 1$	
☞ a. $(--_n)$		-n		-3n
b. $(++_n)$		-n	-1	-3n-1
c. $(-)(++_{(n-1)})$	-1	-(n-1)		-3n-1
d. $(-)(+)_n$	-n			-4n

Charlie O'Hara

(23)   *Privative features can only spread marked value (here +).*

/−+$_n$/	SHARE(F)	MAX(F)	MAX (F)/σ$_1$	H
	$w = 4$	$w = 3$	$w = 1$	
☞ a. $(--_n)$		-n		-3n
☞ b. $(++_n)$		-(n-1)		-3(n-1)
c. $(-)(++_{(n-1)})$	-1	-(n-1)		-3n-1
d. $(-)(+)_n$	-n			-4n

Lombardi (1999) and Baković (2000) note that in the absence of a privileged position controlling the vote, there should be assimilation to the unmarked. Baković (2000) notes that the feature value specific faithfulness constraints, MAX/DEP(F) or IDENT-IO(+F)/(-F) fail to capture these patterns except through stipulation; which faithfulness constraint outranks the other determines the dominant feature, rather than markedness.

   With MAX/DEP(±F) markedness again determines that the unmarked feature value is the dominant one. Let *[−F] be introduced to our system, without its *[+F] counterpart. This constraint assigns a violation mark for each different [−F] feature linked to any (non-zero) number of segments in the output. Presuming there is no relevant positional faithfulness (24), or the positional faithfulness is lower weighted than the *[−F] constraint (25), we get assimilation to [+F] for all segments, even if there is only one [+F] segment in the input, and an unbounded number of [−F] segments.

(24)   *If there is no privileged position, we get assimilation to the unmarked*

/+−$_n$/	SHARE(±F)	MAX(±F)	DEP(±F)	*[−F]	H
	$w = 4$	$w = 3$	$w = 1$	$w = 1$	
☞ a. $(++_n)$		-n			-3n
b. $(--_n)$		-n		-1	-3n-1
c. $(+)(-)_n$	-n			-n	-4n

(25)   *If Positional Faithfulness is low, we get assimilation to the unmarked*

/−$_n$+/	SHARE(±F)	MAX(±F)	*[−F]	MAX(±F)/σ$_1$	H
	$w = 4$	$w = 3$	$w = 2$	$w = 1$	
☞ a. $(++_n)$		-n		-1	-3n-1
b. $(--_n)$		-n		-1	-3n-2
c. $(-)_n(+)$	-n		-n		-4n

In domains with only [-F] segments, if DEP(±F)+MAX(±F) outweighs *[−F], the surface form also has all [−F] segments, either spread as in (26)-b, or not as in (26)-c, depending on whether SHARE outweighs MAX, i.e. whether spreading happens anywhere. If DEP(±F)+MAX(±F) is lower than the markedness constraint, we simply see a language that has no [−F] segments.

(26)   *If no +F is in the word, it remains all −F*

/−−$_n$/	SHARE(±F)	MAX(±F)	DEP(±F)	*[−F]	H
	$w = 4$	$w = 3$	$w = 1$	$w = 1$	
a. $(++_n)$		-(n+1)	-1		-3n-4
☞ b. $(--_n)$		-n		-1	-3n-1
c. $(-)(-)_n$	-n			-(n+1)	-5n-1
d. $(+)(+)_n$	-n	-(n+1)	-(n+1)		-8n-4

5.   Discussion

We've seen that the DEP/MAX(±F) account is largely successful at evading the unbounded asymmetric tradeoffs predicted by many sets of harmony driving and faithfulness constraints. However, plugging this hole seems to spring a leak somewhere else. While this system has been able to make harmony feature spreading, rather than feature changing, it can make some odd predictions in feature changing contexts.

Consider a language where [y] is marked by featural cooccurrence constraint *[-back, +round]. In order to repair the feature value of one [y], say by backing, there must be both deletion of a [-back] feature, and epenthesis of a [+back] feature. However, if multiple [y] underlyingly appear next to each other, rather than epenthesizing one feature for each vowel, one feature could epenthesize and link to all the vowels. This can be caused by an unbounded tradeoff, but is also possible if constraints are defined to avoid one.

This is due to a word-length effect I call *catching up*. This occurs because we see two types of candidates (here faithful, and repairing through harmony), that each violate a constraint (or set of constraints) a number of times proportional to word length. Thus, if faithfulness is weighted above segmental markedness, one syllable words are faithful. But, if constraints that favor the other candidate require at least two syllables but scale in the same way, eventually, but not immediately, these constraints can make up the gap. Thus, faithfulness can be more powerful in short words, getting a head start, but the gang up between segmental markedness and harmony driving eventually catch up, as demonstrated in (27).

(27)   *Catching Up in longer words*

/by/	MAX(±BK)	*y	SHARE(±BACK)	H
	$w = 6$	$w = 3$	$w = 5$	
☞ a. (by)		-1		-3
b. (bu)	-1			-6
/byty/	MAX(±BK)	*y	SHARE(±BACK)	H
☞ c. (by)(ty)		-2	-1	-11
d. (butu)	-2			-12
/bytyly/	MAX(±BK)	*y	SHARE(±BACK)	H
e. (by)(ty)(ly)		-3	-2	-19
☞ f. (butulu)	-3			-18

Charlie O'Hara

Here, if we have less than two adjacent syllables with [y] nuclei, they remain faithful. But, if we have three or more adjacent [y] syllables, they all harmonize and link to a new [+back] feature, in order to avoid the gang effect of *y and SHARE(±BACK)

This happens because the violation score of the faithful candidates scales faster as words get longer than the spreading and changing candidate. In (28), the winning candidate is the one closer to 0. While the faithful candidate starts behind the repair/spread candidate, the magnitude of its harmony score increases faster than the other candidate, because SHARE(±BACK) kicks in at 2 [y]s.

(28)   *Catching Up*
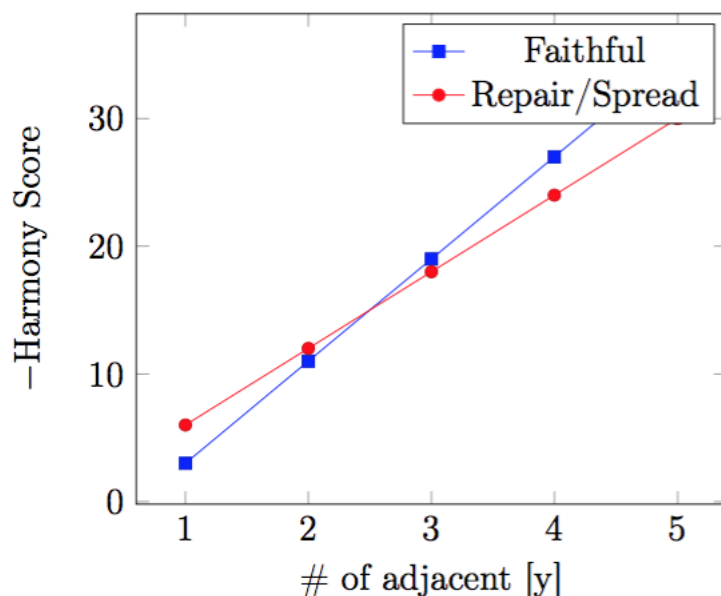
Future work will be needed to evaluate the seriousness of *catching up* pathologies in systems with weighted constraints, and if these can be evaded by reevaluating constraints and representations, or if there is something more intrinsically problematic about them. Assimilation and dissimilation processes likely need constraints that could *catch up*, creating a fundamental problem for systems with weighted constraints. Note that this problem is not restricted to interactions of markedness and faithfulness constraints as presented here, but could also potentially occur with markedness constraints, creating issues for serial Harmonic Grammar as well.

References

Baković, Eric. 2000. Harmony, dominance and control. Doctoral dissertation, Rutgers University, New Brunswick, NJ.

Beckman, Jill N. 1998. Positional faithfulness. Doctoral dissertation, University of Massachusetts Amherst, Amherst.

Causley, Trisha. 1997. Featural correspondence and identity: the Athabaskan case. In *NELS 27*, 93–101.

Finley, Sara. 2008. Formal and cognitive restrictions on vowel harmony. Doctoral dissertation, John Hopkins University, Baltimore, Maryland.

Harrison, K. David. 2000. Topics in the phonology and morphology of Tuvan. Doctoral dissertation, Yale University.

Heinz, Jeffrey, & Regine Lai. 2013. Vowel harmony and subsequentiality. In *Proceedings of the 13th Meeting on the Mathematics of Language*, ed. Andras Kornai & Marco Kuhlmann, 52–63. Sofia, Bulgaria.

Jesney, Karen. 2011. Cumulative constraint interaction in phonological acquisition and typology. Doctoral dissertation, University of Massachusetts Amherst, Amherst.

Jesney, Karen. 2016. Positional constraints in Optimality Theory and Harmonic Grammar. In *Harmonic Grammar and Harmonic Serialism*, ed. John J. McCarthy & Joe Pater. Equinox.

Legendre, Géraldine, Yoshiro Miyata, & Paul Smolensky. 1990. Harmonic Grammar - a formal multi-level conectionist theory of linguistic wellformedness: an application. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, ed. Lawrence Erlbaum, 884–891. Cambridge, MA.

Legendre, Géraldine, Antonella Sorace, & Paul Smolensky. 2006. The optimality theory-harmonic grammar connection. In *The harmonic mind: From neural computation to optimality-theoretic grammar*, ed. Paul Smolensky & Géraldine Legendre, 339–402. MIT Press.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in optimality theory. *Natural Language & Linguistic Theory* 17:267–302.

Lombardi, Linda. 2001. Why place and voice are different: Constraint-specific alternations in optimality theory. In *Segmental phonology in optimality theory: Constraints and representations*, ed. Linda Lombardi. Cambridge University Press.

McCarthy, John J. 2011. Autosegmental spreading in optimality theory. In *Tones and features (clements memorial volume)*, ed. John A. Goldsmith, Elizabeth Hume, & Leo Wetzels. Berlin: Mouton de Gruyter.

McCarthy, John J., & Alan Prince. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers* 18:249–384.

Moreton, Elliott. 1999. Non-computable functions in optimality theory. Ms. University of Massachusetts, Amherst [ROA-667].

Mullin, Kevin. 2011. Strength in harmony systems: Trigger and directional asymmetries. Ms. University of Massachusetts.

Myers, Scott. 1997. OCP effects in optimality theory. *NLLT* 15:847–892.

Pater, Joe. 1999. Austronesian Nasal Substitution and other NC̣ effects. In *The prosody morphology interface*, ed. René Kager, Harry van der Hulst, & Wim Zonneveld, 310–343. Cambridge: Cambridge University Press.

Pater, Joe. 2009a. Review of Smolensky and Legendre (2006). The Harmonic Mind. *Phonology* 26:217–226.

Pater, Joe. 2009b. Weighted constraints in generative linguistics. *Cognitive Science* 33:999–1035.

Pater, Joe. 2016. Universal grammar with weighted constraints. In *Harmonic Grammar and Harmonic Serialism*, ed. John J. McCarthy & Joe Pater. Equinox.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt, & Michael Becker. 2010. Harmonic grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27:77–117.

Prince, Alan, & Paul Smolensky. 1993/2004. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell.

Riggle, Jason. 2004. Generation, recognition, and learning in finite state optimality theory. Doctoral dissertation, University of California Los Angeles.

Rose, Sharon, & Rachel Walker. 2011. Harmony systems. In *Handbook of phonological theory*, ed. Marc van Oostendorp, Colin Ewen, Elizabeth Hume, & Keren Rice. Wiley-Blackwell, 2nd ed edition.

Walker, Rachel. 1997. Faith and markedness in esembi feature transfer [roa-336]. In *Phonology at Santa Cruz*, ed. Rachel Walker, Motoko Katayama, & Daniel Karvonen, volume 5, 103–115.

Charlie O'Hara
charleso@usc.edu