

# Universal Grammar with Weighted Constraints<sup>1</sup>

*Joe Pater, University of Massachusetts, Amherst*

## 1 Introduction

The aim of this paper is to show that weighted constraints hold much promise as the basis of models of Universal Grammar (UG), as in Optimality Theory (OT; Prince and Smolensky 1993/2004). In OT, a universal constraint set is used to analyze individual languages, and to generate predictions about the range of possible languages, that is, about language typology. A ranking of the universal constraint set determines the form of a single language, and every ranking corresponds to some possible language. The view that this framework could benefit from the adoption of weighted rather than ranked constraints is controversial. Prince and Smolensky (1993/2004: 232) claim that this would lead to typologically implausible patterns of constraint interaction – versions of this claim also appear in Prince and Smolensky (1997), Legendre, Sorace and Smolensky (2006) and Smolensky (2006a), and elsewhere. The cited arguments are assessed in sections 2 and 4 below. Weighted constraints were first applied to generative grammar in a theory of gradient syntactic well-formedness named Harmonic Grammar (HG; Legendre, Miyata, and Smolensky 1990; see also Goldsmith 1990, Goldsmith 1993a, and some of the papers collected in Goldsmith 1993b for early applications to phonology). To recognize this innovation, I adopt HG as the name for a version of OT that uses weighted constraints, but as the above citations should make clear, prior work by the original proponents of HG does not take the position that weighted constraints form a plausible basis for OT-style theories of UG.

Part of the promise of HG for UG derives from inherent restrictions on cumulative interaction in this framework. In section 2, building on observations of Prince (2003), I discuss the consequences of what I call the asymmetric trade-off requirement on gang effects. Only certain patterns of constraint violation can produce gang effects, and hence contain the potential for OT-HG differences. To further demonstrate this inherent restrictiveness, I contrast HG with Smolensky's (2006b) Optimality Theory with Local Constraint Conjunction (OT-LC), which does not impose similar restrictions on cumulative constraint interaction, and thus generates a set of implausible typological predictions that are not shared by HG. I further demonstrate the restrictiveness of HG by pointing out that meeting the asymmetric trade-off requirement is not a guarantee of an OT-HG difference: some gang effects are vacuous, in that they yield no differences in the predictions of the two frameworks.

A second part of the promise of weighted constraint theories of Universal Grammar derives from the ability of HG to generate attested patterns that fall out of the reach of OT using the same set of constraints. That is, HG permits new theories of Con, the universal constraint set. In section 3, I discuss the compatibility of scalar constraints with HG (building on Flemming 2001), and their incompatibility with OT (building on Prince and Smolensky 1993/2004, McCarthy 2003). Because a fleshed-out HG theory of any domain will likely be operating with a different constraint set than an OT one, the assessment of the relative success of OT and HG as general frameworks for typological study becomes more complicated, and more interesting.

A further complication in making OT-HG comparisons arises when we take into account the limits on patterns of constraint interaction that follow when Prince and Smolensky's (1993/2004) parallel theory of candidate generation and evaluation is replaced by a serial one (see McCarthy this volume for an introduction to Harmonic Serialism and an overview of

results). In section 4, I show that some clearly undesirable predictions of HG are eliminated when only a single application of an operation is allowed in constructing a candidate. The upshot of this discussion is that some pathological predictions of parallel HG are plausibly due to the globality of parallel evaluation, rather than to the power of weighted constraint interaction.

Typological research using weighted constraints may well have been hindered in the past by the relative difficulty of finding by hand a correct set of constraint weights for a given set of linguistic data, and especially of determining what all of the languages are that a given set of constraints can generate. This paper introduces the basic techniques for doing both of these analytic tasks, and discusses how they can be aided by using computational techniques, in particular the implementation of Potts *et al*'s (2010) Linear Programming methods in OT-Help (Staub's *et al.* 2010). A great deal of the prior and ongoing computational work on HG focuses on its use in learning with probabilistic versions of the framework (Maximum Entropy Grammar, Goldwater and Johnson 2003, Noisy HG, Boersma and Pater this volume). Because the goal here is to examine the differences between the typological predictions of standard OT and a minimally different version with weighted constraints, I adopt a categorical version of the theory here. Section 5 briefly discusses the translation of the present model and the results obtained with it to probabilistic variants of HG.

## 2 Asymmetric trade-offs

### 2.1. Background

We start with the simple HG tableau in (1), which has an input with a pair of voiced obstruents, and as output candidates the result of changing the voicing of either one, of both, or of neither. The optimum is the candidate with the highest *Harmony*, which in HG is the weighted sum of constraint violations. Constraint violations are indicated with negative integers, and the constraint weights are given immediately beneath the constraint names: 3 for \*Coda-Voice, and 2 for Ident-Voice. The Harmony score for each candidate is given in the rightmost column. In this tableau, the candidate with final devoicing, [bat], receives a score of  $-1$  on Ident-Voice, which penalizes each input consonant whose voice specification is changed in the output (McCarthy and Prince 1999). Since Ident-Voice has a weight of 2, [bat] has Harmony of  $-2$ . This candidate has the highest score, and is thus optimal. In particular, the faithful [bad], which violates a constraint against voiced codas, \*Coda-Voice, gets a lower Harmony score ( $-3$ ) because that constraint's weight is higher.

#### (1) *Final devoicing in HG*

/bad/	*CODA-VOICE 3	IDENT-VOICE 2	
a. bad	-1		-3
b. pad	-1	-1	-5
c. → bat		-1	-2
d. pat		-2	-4

The weights in the above tableau are partially arbitrary; there is an infinite set of weights that could be used to make [bat] optimal – e.g weights (30, 20) and (0.111, 0.110) instead of (3, 2)

would produce the same optima. The non-arbitrary aspect is that for [bat] to be optimal, the weight of \*Coda-Voice must be greater than that of Ident-Voice. With the reverse relationship, [bad] becomes optimal.

More generally, given a desired optimum, or “Winner”, a set of failed candidates, or “Losers”, and their associated vectors of scores – which can also be positive rewards – on a set of constraints, we can produce a set of linear inequalities, or *weighting conditions*, that must obtain if the Winner is to be made correctly optimal (like Prince’s 2002 OT *ranking conditions*). We can obtain a useful representation for examining weighting conditions by subtracting the (unweighted) scores of the Winner from those of a Loser, thus producing an HG comparative vector (see again Prince 2002 in OT; in HG see Goldwater and Johnson 2003, Potts *et al.* 2010, Bane and Riggle 2012, and Boersma and Pater this volume). The comparative vectors from the candidates in (1), with [bat] as the Winner, are shown in (2).

(2) *Comparative vectors for (1)*

W ~ L	*CODA-VOICE	IDENT-VOICE
a. [bat] ~ [bad]	+ 1	–1
b. [bat] ~ [pat]		+ 1
c. [bat] ~ [pad]	+ 1	

The weighting conditions derive from the requirement that for the Winner to be correctly optimal, the weighted sum of the scores in each row must be greater than zero. For example, for row (2a.), we have the requirement that 1 times the weight of \*Coda-Voice, plus –1 times the weight of Ident-Voice, must be above zero, which is satisfied by the weights from (1) – *i.e.*  $1 \times 3 + -1 \times 2 = 1$ . When all of the non-zero scores are +1 and –1, as in this row and all others in (2), we can also simply say that the sum of the weights of the constraints preferring the Winner must be greater than the sum of the weights preferring the Loser.

Our comparative vectors have not introduced any new weighting conditions beyond the one we noted before: as indicated by vector (2a.), the weight of \*Coda-Voice must be greater than that of Ident-Voice. So long as weights are limited to positive values, it is impossible to make either [pat] or [pad] beat [bat]; any positive weights applied to the vectors in (2b.) and (2c.) will lead to weighted sums above zero. These are instances of simple harmonic bounding, in that [pat] and [pad] are both harmonically bounded by [bat] because they each have a proper superset of its constraint violations. As Prince (2003) points out, positively weighted constraints preserve simple harmonic bounding relations from OT.<sup>2</sup> Zero weights could make a simply harmonically bounded candidate tie for optimality, but could not make it the sole optimum. Negative weights can make a harmonically bounded candidate solely optimal in its tableaux, and thus need to be banned in a version of HG that aims to function anything like OT.

With more candidates and more tableaux, the set of weighting conditions can become more complex, making it difficult to find a correct set of weights by hand. Weighting conditions are linear inequalities, and a system of linear inequalities can be solved by Linear Programming’s simplex algorithm. Potts *et al.* (2009) show how to translate HG learning problems into systems solvable by the simplex, essentially by making the same transition from tableaux to weighting conditions that we have made here. Not only can the simplex be used to find a set of weights that meets a set of weighting conditions of arbitrary complexity, but it can also detect when no correct weighting exists because the weighting conditions are inconsistent. As such, it does for

HG what Recursive Constraint Demotion (RCD, Tesar and Smolensky 2000) does for OT. The inconsistency detection property of RCD makes it particularly useful in calculating the set of languages produced by a set of constraints, that is, the sets of candidates across tableaux that can be made jointly optimal. It was first applied in this way in the typology calculator implemented in OT-Soft (Hayes, Tesar and Zuraw 2003). Staubs *et al.*'s (2010) OT-Help uses Potts *et al.*'s simplex application, as well as RCD, to calculate and compare OT and HG typologies.

It is important to note that both RCD and the Linear Programming application in Potts *et al.* (2009) only work for cases in which there is a single optimum per candidate set. Because they rely on these algorithms to calculate typologies, OT-Soft and OT-Help will not find languages with tied optima. A limit to single optima (for non-identical candidates) is in fact a useful idealization for a comparison between the standard version of OT and a minimally different version with weighted constraints, since the standard version of OT generally produces only single optima. Throughout this paper I discuss only languages with one optimum per tableau, but return to this issue of ties in the discussion of probabilistic versions of HG in section 5.

## 2.2. Loanword devoicing in Japanese as cumulative constraint interaction

To see how weighted constraints can produce results that diverge from ranked ones, we can consider a slightly more complicated example in the phonology of obstruent voicing. In Japanese, only a single voiced obstruent is usually permitted in a word (see Ito & Mester, 1986, 2003). This restriction is termed Lyman's Law (Lyman, 1894). In loanwords, however, multiple voiced obstruents are permitted (Kawahara, 2006, 2011; Nishimura, 2003, 2006; all data are from Kawahara 2006):

### (3) Violations of Lyman's Law in loanwords

[bagi:] 'buggy'	[bogi:] 'bogey'	[bobu] 'Bob'
[bagu] 'bug'	[dagu] 'Doug'	[giga] 'giga'

Japanese also has a restriction against obstruent voicing in geminates. But again, in loanwords, voiced geminate obstruents are permitted:

### (4) Voiced/voiceless obstruent geminate near-minimal pairs in Japanese loanwords

[web:u] 'web'	[wip:u] 'whipped (cream)'
[sunob:u] 'snob'	[sutop:u] 'stop'
[hab:uru] 'Hubble'	[kap:uru] 'couple'
[kid:o] 'kid'	[kit:o] 'kit'
[red:o] 'red'	[autoret:o] 'outlet'
[hed:o] 'head'	[met:o] 'helmet'

Devoicing occurs just when a loanword contains both a voiced geminate and another voiced obstruent. The geminate is optionally, but categorically, devoiced:

### (5) Optional devoicing of a geminate in Lyman's Law environment

[gud:o] ~ [gut:o] 'good'	[dog:u] ~ [dok:u] 'dog'
[bed:o] ~ [bet:o] 'bed'	[deibid:o] ~ [deibit:o] 'David'
[dored:o] ~ [doret:o] 'dredlocks'	[bag:u] ~ [bak:u] 'bag'
[bad:o] ~ [bat:o] 'bad'	[bud:a] ~ [but:a] 'Buddha'

[dorag:u] ~ [dorak:u]‘drug’      [big:u] ~ [bik:u]‘big’

According to Nishimura (2003, 2006) and Kawahara (2006, 2011), such devoicing is judged unacceptable (or much less acceptable; Kawahara 2011) in the word types illustrated in both (3) and (4).

In HG, this devoicing pattern can be analyzed as being due to two independently motivated constraints. This analysis draws on Nishimura’s (2003, 2006) account using Smolensky’s (2006b) OT with Local Conjunction; the possibility of an HG reanalysis was suggested by Shigeto Kawahara (p.c.). The first constraint, \*Vce-Gem expresses a cross-linguistically common ban against voiced obstruent geminates, which can be held responsible for their absence in native Japanese words. The other constraint is OCP-Voice, which Ito and Mester (1986) propose to account for the Lyman’s Law restriction in native Japanese words. This constraint penalizes every sequence of voiced obstruents, even ones separated by any number of segments (other than voiced obstruents). For example, [deibit:o] and [dored:o] each have one violation of OCP-Voice ([d...d:]), and [deibid:o] has two ([d...b], [b...d:]), while [deipit:o] and [doret:o] have none. The weight of Ident-Voice is greater than that of each of OCP-Voice and \*Voice-Obs, so that a word can have either a pair of voiced obstruents, as in (6), or a voiced geminate as in (7).

(6) Multiple obstruents permitted

/bobu/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
	3	2	2	
a. → bobu		-1		-2
b. bopu	-1			-3
c. pobu	-1			-3

(7) Voiced geminates permitted

/web:u/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
	3	2	2	
a. → web:u			-1	-2
b. wep:u	-1			-3

In the tableau in (8), we have a word with both a voiced geminate and another voiced obstruent. As the summed weight of OCP-Voice and \*Vce-Gem is greater than that of Ident-Voice, the geminate devoices (the probabilistic versions of HG discussed in section 5 could generate the observed optionality; see Pater 2009a for an explicit analysis).

## (8) Devoicing of geminate in the context of another voiced obstruent

/dog:u/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
	3	2	2	
a. dog:u		-1	-1	-4
b. → dok:u	-1			-3
c. tog:u	-1		-1	-5
d. tok:u	-2			-3

In this *gang effect*, we have two constraints ganging up to overcome a third one with higher weight. Due to strict domination, OT cannot express this sort of constraint interaction. In this case, if Ident-Voice were ranked above each of \*Vce-Gem and OCP-Voice, as required for the first two tableaux, the optimal output for /dog:u/ would be [dog:u], rather than [dok:u].

The OT-HG difference can be clearly seen in the comparative vectors in (9). For HG, the requirements that the weight of each of OCP-Voice and \*Voice-Gem be lower than that of Ident-Voice (9a. and 9b.) do not contradict the requirement that their summed weight be greater (9c.). In OT, on the other hand, (9c.) requires that either OCP-Voice or \*Vce-Gem dominate Ident-Voice, which would contradict one of (9a.) or (9b.).

## (9) Comparative vectors for Japanese loanword devoicing

W ~ L	OCP-VOICE	*VOICE-GEM	IDENT-VOICE
a. [bobu] ~ [bopu]	-1		+1
b. [web:u] ~ [wep:u]		-1	+1
c. [dok:u] ~ [dog:u]	+1	+1	-1

Many other linguistic patterns have been analyzed in terms of weighted constraint interaction (in phonology, see e.g., Guy, 1997 as a precedent, and more recently Farris-Trimble 2008, Potts *et al.* 2010, Jesney and Tessier 2011, and Jesney 2012/this volume). This example provides particularly striking evidence for HG because OCP-Voice and \*Vce-Gem are independently motivated in the phonology of Japanese. As noted above, the OCP-Voice constraint was posited by Ito and Mester (1986) to account for the fact that the native vocabulary is restricted by Lyman's Law, and for the morphological alternations that show its effects. In generative linguistics outside of OT and HG, an active constraint is a true statement about the domain in which it applies – in other words, an active constraint is inviolable. Given the examples of loanwords with pairs of voiced obstruents like [bobu], OCP-Voice would have to be considered inactive, at least for loanwords (see Ito and Mester, 1995). The prediction, then, is that OCP-Voice should not have any effect on the devoicing of geminates. Similarly, if OCP-Voice is outranked by Ident-Voice for the OT account of [bobu]-type words, then as we have seen, it cannot participate in the devoicing of geminates. Nishimura's (2003, 2006) discovery of the cumulative effect of OCP-Voice and \*Vce-Gem in Japanese loanwords thus falsifies these predictions of frameworks with inviolable or ranked constraints, and counts in favor of HG's weighted ones.

The predictions of these frameworks depend, of course, on the contents of the constraint sets. With a different constraint set, one could analyze the Japanese case with either inviolable

constraints or ranked ones. Nishimura (2003, 2006) posits a conjoined version of OCP-Voice and \*Vce-Gem, which is violated if and only if a word contains violations of both constraints. In the following section, we will turn to a discussion of the problematic typological predictions made by OT with constraint conjunction. Less obviously problematic is Kawahara's (2006) introduction of a singleton-specific Ident-Voice, which fails to apply to geminates, and can thus rank above OCP-Voice to protect singletons in the Lyman's Law environment, while still allowing devoicing of geminates in that context. The point, though, is that the cumulative interaction was *predicted* only by HG.

### 2.3. The necessity of asymmetric trade-offs

Japanese loanword devoicing involves an *asymmetric trade-off*: the choice between [dok:u] and [dog:u] in (8) trades a single Ident-Voice violation against violations of OCP-Voice and \*Vce-Gem. Because asymmetric trade-offs are required to produce gang effects, in their absence HG produces the same typology as OT.

We will now look at this key restrictiveness result in some detail. We can call a constraint that distinguishes an optimum from another candidate – that is, on which the optimum and some other candidate have different violation scores – a distinguishing constraint. A gang effect can be generally defined simply as an HG tableau in which the distinguishing constraint with highest weight<sup>3</sup> does not prefer the optimal candidate – e.g. Ident-Voice does not prefer the optimal [dok:u] in (8). If we convert the weight values to OT ranks, the resulting OT hierarchy will choose a different optimum than the HG weighting did – e.g. OT chooses [dog:u], HG chooses [dok:u]. This is a “gang” effect because the lower-valued constraint violations are ganging up to overcome the higher weighted one.

The asymmetric trade-off requirement falls out directly from this definition. Let us assume that the HG optimum in this scenario has just one more violation mark on the highest weighted distinguishing constraint than the OT optimum – (8) again serves as an example, since HG optimal [dok:u] violates just Ident-Voice, which the OT optimum [dog:u] satisfies. For the HG optimum to in fact be optimal, it must be then be the case that the OT optimum incurs at least two constraint violations that are not shared by the HG optimum: at least two unshared violations – OCP-Voice and \*Vce-Gem in our example – are required for their summed weight to be greater than that of the highest weighted distinguishing constraint (e.g. Ident-Voice). More generally, in a gang effect, if the difference between the OT and the HG optima on the highest weighted distinguishing constraint is  $n$ , then there must be at least  $n + 1$  constraint violations incurred only by the OT optimum. Therefore, a set of violation profiles that can produce a gang effect must have a pair of candidates in which  $n$  violations of one constraint trade against at least  $n + 1$  violations of some other constraint(s), that is, they must contain an asymmetric trade-off.

The candidate set in (10) demonstrates a situation in which the asymmetric trade-off requirement is not met. For every violation of \*Coda-Voice that is taken away, one violation of Ident-Voice is added, and vice versa. Keeping aside the possibility of tied optima in both theories, no set of weights will produce an OT-HG difference: if \*Coda-Voice has a higher weight than Ident-Voice, all underlying voiced consonants in surface coda position undergo devoicing, and with the reverse weighting, they will all surface with voicing intact. This is just the same as the effect of ranking the constraints with respect to one another.

## (10) Symmetric trade-off

/dagbad/	*CODA-VOICE	IDENT-VOICE
a. dag.bad	-2	
b. dak.bad	-1	-1
c. dag.bat	-1	-1
d. dak.bat		-2

The equivalence between HG and OT when constraint violations trade one-to-one was first noted by Prince (2003), who points out that in this situation, the result of the “greater than” relation between positive weights in HG is the same as the “dominates” relation in OT – any set of weights that meets these inequalities will behave just like the ranked constraints (i.e. “Anything Goes”).

Prince (2003) is concerned only with the issue of translating between ranking and weighting, that is, of finding a set of weights that will behave as an OT hierarchy. Here we see that there is also a typological consequence of this observation. This restrictiveness result is far from trivial, and is in some ways non-intuitive. This is highlighted by the fact that at least one of the differences between ranking and weighting that Prince and Smolensky (1997) use to motivate OT does not in fact exist, due to the asymmetric trade-off requirement.

In making the case for OT to an interdisciplinary audience, Prince and Smolensky (1997: 1604) draw the generalization that:

- (11) In a variety of clear cases where there is a strength asymmetry between two conflicting constraints, no amount of success on the weaker constraint can compensate for failure on the stronger one.

They attribute this type of phenomenon to strict domination property of ranked constraints. As an example, they discuss the interaction of NoCoda and Parse. Parse is a faithfulness constraint that demands that input segments be parsed into output syllable structure; a consonant that is unparsed is unpronounced ( $\approx$  deleted). Prince and Smolensky (1997: 1606) state that:

- (12) Domination is clearly “strict” in these examples: No matter how many consonant clusters appear in an input, and no matter how many consonants appear in any cluster, [the grammar with NoCoda  $\gg$  Parse] ... will demand that they all be simplified by deletion (violating Parse as much as is required to eliminate the occasion for syllable codas), and [the grammar with Parse  $\gg$  NoCoda] ... will demand that they all be syllabified (violating NoCoda as much as is necessary). No amount of failure on the violated constraints is rejected as excessive, as long as failure serves the cause of obtaining success on the dominating constraint.

In this passage, Prince and Smolensky offer as one illustration of strict domination the observation that the number of consonant clusters, that is, of potential codas, does not affect whether deletion occurs or not – see section 4 on the second case, the number of consonants in a single cluster. The table in (13) shows that the trade-offs across candidates’ violation profiles have the same symmetry as coda devoicing in (9). Unparsed segments are placed between angled brackets; as they are unparsed, they do not violate NoCoda.

## (13) Symmetric trade-off

/dagbadga/	NoCODA	PARSE
a. dag.bad.ga	-2	
b. da<g>.bad.ga	-1	-1
c. dag.ba<d>.ga	-1	-1
d. da<g>.ba<d>.ga		-2

Regardless of the number of potential codas, there are only two outcomes: if Parse has a greater weight than NoCoda, the language will permit codas (as in [dag.bad.ga]), and if the relationship is reversed, all potential codas will fail to be pronounced (as in [da<g>.ba<d>.ga]). Thus, strict domination is irrelevant to the irrelevance of the number of potential codas; the lack of that number's effect on whether deletion applies is a prediction of an optimization system with these constraints, be they ranked or weighted.

I can also point to some anecdotal evidence that this HG restrictiveness result is non-intuitive. When I first started investigating the relationship between weighted and ranked constraints, many of the cases that I imagined would yield OT-HG differences turned out to involve symmetric trade-offs and thus failed to distinguish the frameworks. Such cases have also sometimes been offered as potential instances of OT-HG differences by participants in courses and audience members at talks in which I have presented this material. I suspect that the source of this misleading intuition is the failure to appreciate the role of optimization in affecting what HG can and cannot do. This version of HG, like OT, is choosing the best outcome in a candidate set, rather than simply imposing a numerical cut-off on degree of violation (though see section 3 and 4 on situations in which it can do something like this). There is no lower bound on the Harmony of an output – any amount of ill-formedness will be tolerated if the cost of avoiding it is higher.

## 2.4. The relative restrictiveness of HG and OT with Local Conjunction

This section illustrates the force of the asymmetric trade-off requirement in HG through a comparison with Smolensky's (2006b) OT with locally conjoined constraints (henceforth OT-LC), a framework that does not require asymmetric trade-offs in order to generate effects of cumulative violation. This comparison builds on that of Legendre, Sorace and Smolensky (2006), though does come to the opposite conclusion about the relative merits of OT-LC and HG as frameworks for Universal Grammar.

Smolensky (2006b: 43) defines local conjunction as in (14). This operation of conjunction yields a new constraint that is separately rankable in the constraint hierarchy.

(14) Local conjunction within a domain  $D$ 

\*A& <sub>$D$</sub> \*B is violated if and only if a violation of \*A and a (distinct) violation of \*B both occur withing a single domain of type  $D$ .

The original motivation for OT-LC, and the one that Smolensky (2006b) focuses on, is the reduction of complex markedness constraints to more basic primitives. Following Ito and Mester (2003: 26), we can take as an example \*CodaVoice. This can be expressed as the conjunction of two independently needed constraints, NoCoda and \*VoiceObs ('assign a violation mark to a voiced obstruent'). When these are conjoined in the domain of a segment, the resulting constraint penalizes voiced obstruent codas, but neither voiceless codas nor voiced

obstruents. If the conjoined constraint is ranked above Ident-Voice, which is in turn ranked above \*VoiceObs, the result is coda devoicing (the position of NoCoda is arbitrary):

(15) Local conjunction analysis of final devoicing

/bad/	NoCODA& <sub>SEG</sub> *VOICEOBS	IDENT-VOICE	*VOICEOBS	NoCODA
a. bad	-1		-2	-1
b. pad	-1	-1	-1	-1
c. → bat		-1	-1	-1
d. pat		-2		-1

Because a single violation of Ident-Voice cannot be used to avoid violations of both NoCoda and \*VoiceObs, with just the unconjoined constraints in (15), HG cannot generate this pattern. This is shown clearly by the comparative vectors for [bat] as a winner in (16). The shared NoCoda violations cancel out, and the remaining violation pattern is a one-to-one trade between \*VoiceObs and Ident-Voice. No set of positive weights will pick [bat] as optimal.

(16) Comparative vectors for unconjoined constraints in (15)

W ~ L	*VOICEOBS	IDENT-VOICE	NoCODA
a. [bat] ~ [bad]	+1	-1	
b. [bat] ~ [pat]	-1	+1	
c. [bat] ~ [pad]			

Legendre, Sorace and Smolensky (2006) point to similar cases of HG not being able to reduce a constraint to the activity of more basic ones. One example they give is that HG cannot ban [x] with general constraints against velars and fricatives, when the inventory contains both velars and non-velar fricatives, and in which constraint violations trading off against \*Velar and \*Fricative are non-overlapping (e.g. Ident-Place and Ident-Continuant). They argue that OT-LC's success in this regard counts in its favor.

In my view, OT-LC's success on cases like these does not clearly pick it over HG and standard OT. Whether complex markedness constraints like \*CodaVoice and \*[x] should be reduced to more basic constraints is a matter of some general controversy. The idea that \*CodaVoice is the sum of the effects of a constraint against codas and one against voiced obstruents harks back in some ways to theories of prosodic licensing (e.g. Ito 1986, Goldsmith 1990, Lombardi 1991, Steriade 1995), which see contextual markedness as the inability of marked prosodic contexts to license marked segments (see Ito and Mester 2003: 28-29 on this connection). Much work in OT has questioned this approach, analyzing contextual markedness as the effect of rather specific, substantively motivated constraints (e.g. Pater 1999, Steriade 1999). The empirical motivation for these alternatives is that contextual markedness displays asymmetries that are not captured by prosodic licensing: the set of marked contexts is not the same for every marked segment, and markedness relationships between segments can be reversed across contexts (see Barnes 2006 for an extensive recent critique of prosodic licensing theory). Like prosodic licensing, this reductionist application of OT-LC also fails to express these asymmetries. Whether building a phonological theory with very specific phonetically grounded universal constraints is the right response to such asymmetries is of course also a

matter of controversy (see Blevins 2004 and Hayes, Kirchner and Steriade 2004 for two poles of the debate, and Pater and Moreton 2012 for an intermediate position). However, the greater reductionism possible in an OT-LC account of contextual markedness does not seem to be a knockdown argument for it over OT and HG, especially since that reductionism is only obtained in a somewhat abstract sense, given the presence in the grammar of the conjoined constraint, and the absence of a learning mechanism that yields conjunction.

Legendre, Sorace and Smolensky (2006) point out that in comparison with HG's linear model (the harmony function is a linear equation), OT-LC is a *superlinear* theory of constraint interaction. In linguistic theory, greater power is of course a double-edged sword. The superlinearity of OT-LC also allows it to generate unattested linguistic patterns not generated by HG. One example, discussed first by Ito and Mester (1998), can be produced by conjoining NoCoda with Ident-Voice in the domain of the segment (thanks to Matt Wolf for bringing this to my attention).

(17) Local conjunction analysis of initial devoicing

	/bad/	NoCODA& <sub>SEG</sub> IDENT-VOICE	*VOICEOBS	IDENT-VOICE	NoCODA
a.	bad		-2		-1
b.	→ pad		-1	-1	-1
c.	bat	-1	-1	-1	-1
d.	pat	-1		-2	-1

The ranking \*VoicedObs >> Ident-Voice leads to devoicing, which is blocked by the dominant conjoined constraint when it occurs in coda position, where it violates both NoCoda and Ident-Voice. This pattern is referred to as a markedness reversal, since the feature is now protected in only the marked environment (see Lubowicz 2005 for further cases). NoCoda and Ident-Voice cannot interact in this way in HG because the NoCoda violation is shared by all of the candidates, and is hence irrelevant to the outcome.

The fact that [pad] can be the optimal output for /bad/ in OT-LC but not in HG points to an important difference between the frameworks. As discussed in section 2.1 above, simple harmonic bounding relations from OT are preserved in a version of HG with positive weights. This example shows that they are not maintained under constraint conjunction. Critiques of OT-LC have generally focused on its excessive power (see e.g. McCarthy 1999, 2003, Padgett 2002), and have brought out two main issues: that it does not require cumulative interactions to be *local*, or *co-relevant*. We will now see that locality and co-relevance restrictions emerge from the structure of HG, due to the asymmetric trade-off requirement.

An example of the locality problem for OT-LC arises when the domain of the conjoined constraint used for coda devoicing is widened from the segment to the word. The tableaux in (18) and (19) illustrate the result: a consonant in any position will devoice if there is a coda anywhere in the word.

(18) Onset consonant devoices when the word contains a coda

/balatak/	NoCODA& <sub>wd</sub> *VOICEOBS	IDENT-VOICE	*VOICEOBS	NoCODA
a. balatak	-1		-1	-1
b. → palatak		-1		-1

(19) Onset consonant is voiced when the word has no coda

/balatak/	NoCODA& <sub>wd</sub> *VOICEOBS	IDENT-VOICE	*VOICEOBS	NoCODA
a. → balatak			-1	-1
b. palatak		-1		-1

HG cannot replicate this result, again because the violations of NoCoda and \*VoiceObs cannot both be removed with a change in voicing.

The co-relevance problem refers to the ability of OT-LC to express cumulative interaction between constraints that are in fact independent of one another. We can construct an example by conjoining a constraint demanding place agreement (Agree-Place) with \*Coda-Voice (which itself could be the conjunction in (15)). The tableau in (20) shows that when Ident-Voice is ranked beneath the conjoined constraint but above \*Coda-Voice, consonants only devoice in the context of a following heterorganic consonant.

(20) Coda devoicing in the presence of place disagreement

/wadmad/	*CODA-VOICE& <sub>SEG</sub> AGREE-PLACE	IDENT-VOICE	*CODA-VOICE	AGREE-PLACE
a. wadmad	-1		-2	-1
b. → watmad		-1	-1	-1
c. watmat		-2		-1

Although coda devoicing and place agreement are common patterns cross-linguistically, they do not interact in this way. Once again, this pattern cannot be expressed in HG: Ident-Voice violations trade with \*Coda-Voice violations, not with Agree-Place.

Bakovic (2000) and Lubowicz (2005) propose restrictions on OT-LC that are aimed at addressing the co-relevance problem. Tellingly, the restriction that Lubowicz (2005) imposes on OT-LC cumulative interactions emerges from the nature of HG cumulativity: for two markedness constraints to have a cumulative effect, they must be satisfied by violating a single other (faithfulness) constraint. The relative merits of OT-LC and HG remain to be fully assessed, and this comparison is complicated by the fact that it depends on what limitations one applies to conjunction. For example, OT-LC with self-conjoined constraints forming a “Power Hierarchy” (Smolensky 2006b) can generate the same counting effects in stress placement that Legendre, Sorace and Smolensky (2006) present as fatal for HG (though see section 4). Paul Smolensky (p.c.) suggests that one might entertain a version of OT-LC that operates without this sort of self-conjunction. It is important too to note in this context that it is not the case that the full version of OT-LC is in a superset relation with HG in terms of the languages it generates. As McCarthy (2002) points out, HG does not require constraints in a gang effect to share a domain, and it is possible to construct patterns that only HG generates (for example, on

the basis of cumulative effects of Max and markedness constraints, which Moreton and Smolensky 2000 claim are impossible in OT-LC). However, the inherent locality and co-relevance properties of HG, which derive from the asymmetric trade-off requirement, restrict HG in such a way that this freedom does not appear to lead to anything like the kind of overgeneration that arises from superlinear OT-LC.

## 2.5. The insufficiency of asymmetric trade-offs

In order to provide a reasonably accurate general depiction of the relationship between the patterns produced by weighted and ranked constraint interaction, it is important to emphasize that while an asymmetric trade-off is a necessary condition for an HG–OT difference, it is not a sufficient one. We get a simple example by adding \*VoiceObs to the constraint set used for coda devoicing in (1).

### (21) Asymmetric trade-off with voicing constraints

/bad/	*CODA-VOICE	*VOICEOBS	IDENT-VOICE
a. bad	-1	-2	
b. bat		-1	-1
c. pad	-1	-1	-1
d. pat			-2

Here we have an asymmetric trade-off between the violations of \*CodaVoice and \*VoiceObs in [bad], and the violation of Ident-Voice in [bat]. The shared violation of \*Voice cancels out, resulting in the same two-to-one trade that we saw in Japanese; the comparative vector in (22a.) looks just like the [dok:u] ~ [dog:u] vector in (9c.).

### (22) Comparative vectors for coda devoicing

W ~ L	*CODA-VOICE	*VOICEOBS	IDENT-VOICE
a. [bat] ~ [bad]	+1	+1	-1
b. [bat] ~ [pat]		-1	+1
c. [bat] ~ [pad]	+1		

By adding \*VoiceObs to the constraint set, we have changed the weighting conditions (compare the vectors in (2), which have only the leftmost two constraints of (22)): Ident-Voice now must have a greater weight than \*VoiceObs (22b.), and the sum of \*Coda-Voice and \*VoiceObs must be greater than that of Ident-Voice.

We have not, however, introduced a difference in the typologies produced by OT and HG: both theories can make [bad], [bat] and [pat] optimal, but not [pad]. \*Coda-Voice and \*Voice can participate in a gang effect to make [bat] win, but that gang effect would be vacuous in that it does not lead to an OT-HG difference.

OT can deal with the comparative vectors in (22) because there are no data that force \*Coda-Voice to be dominated by Ident-Voice: \*Coda-Voice prefers only Winners. More generally, it is impossible to create a dataset in which the sum of \*Coda-Voice and \*Voice are required to be greater than Ident-Voice, but in which \*Coda-Voice alone must have a lower weight, because every instance of a \*Coda-Voice violation is also a \*Voice violation. Many

constraints in the OT literature stand in the specific-to-general relationship exemplified by \*Voice and \*Coda-Voice, in which the specific constraint assigns violations to a proper subset of the forms violated by the general constraint (see e.g. de Lacy 2006). The gang effect between any of these pairs of constraints will usually be vacuous, indistinguishable from the pattern produced by obedience to the specific constraint alone (though see Pater 2010). In most of these cases, then, asymmetric trade-offs will fail to yield HG–OT differences in typology (though see Jesney and Tessier 2011 on advantages of HG for modeling language learning when faithfulness constraints are in a specific-to-general relationship).

As another type of example of a vacuous gang effect, we turn to a case of constraint interaction that Prince and Smolensky (1993/2004) point to as illustrating the difference between ranking and weighting. It involves the pair of candidates in their tableau (183A), which forms part of their analysis of Lardil final vowel truncation (Hale, 1973):

(23) Tableau 183A from Prince and Smolensky (1993/2004)

/yiliyili/	FREE-V	ALIGN	PARSE	NoCODA
a. → yi.li.yil.<i>		–1	–1	–1
b. yi.li.yi.li	–1			

Parse and NoCoda are violated in the first candidate because the final vowel is unparsed and the last consonant is syllabified as a coda. These violations serve to satisfy Free-V, which demands that the word-final vowel be unparsed. Satisfaction of Free-V also forces a violation of the constraint Align, which requires the edge of the word to coincide with a syllable boundary. Prince and Smolensky make two comments about this tableau. The first is on p. 144:

(24) The relative harmonies of .yi.li.yil.<i> (183 A.i) and .yi.li.yi.li. (183 A.ii) pointedly illustrate the strictness of strict domination. Fully parsed .yi.li.yi.li. is less harmonic than truncated .yi.li.yil.<i> even though it violates only one constraint, while the truncated form violates three of the four lower ranked constraints...

The second is on p. 148:

(25) *Strictness of strict domination.* In several examples the correct analysis violates many constraints, and its optimality rests crucially on the fact that competitors with a cleaner record overall happen to violate some single dominant constraint. Recall the discussion of /yiliyili/ in 7.3.2: a strong contender violating just one constraint is bested by an optimal parse violating three of the four less dominant constraints. This effect highlights the content of the central evaluative hypothesis, and sets the theory apart from others in which richer notions of ‘weighting’ and ‘trade-off’ are entertained.

It is in fact not clear how the Lardil example is meant to set ranking apart from weighting. First, we can obviously assign a set of weights to the constraints to pick the correct optimum: for any finite set of data, any set of OT optima can also be made optimal in HG (Prince and Smolensky 1993/2004). So long as the weight of Free-V is greater than the summed weights of Align, Parse, and NoCoda, [yi.li.yil.<i>] will emerge as optimal in an HG version of (23).

Less obviously, any gang effect between the three constraints violated by [yi.li.yil.<i>] would be vacuous. This is due to another kind of specific-to-general relationship that obtains between Free-V and both Align and Parse: Any candidate that satisfies Free-V necessarily violates Align and Parse, but not vice versa (Prince & Smolensky, 1993/2004: 7.2.1). Because

Free-V satisfaction entails the violation of these constraints, a gang effect that involves Align and/or Parse with NoCoda in blocking deletion would be vacuous: The sum of the effects of Align and/or Parse with NoCoda in forcing the violation of Free-V would be the same as the effect of NoCoda alone. These constraints fail to provide a one-to-one trade-off between NoCoda and Free-V. In the absence of this one-to-one trade-off, there would be no occasion for the lower weight of NoCoda than Free-V to show its effect. As this gang effect is vacuous, it does not produce a divergence between the typological predictions of HG and OT.

### 3 OT and HG with different scalar constraint sets

In the last section, I mentioned Prince and Smolensky's (1993/2004) observation that HG is in a superset relation with OT: any finite set of OT optima can also be made optimal in HG (p. 236), but some HG optima cannot be OT optima (p. 233).<sup>4</sup> One might conclude from this that OT is inherently more restrictive in terms of its typological predictions. However, this follows only if OT and HG share the same constraint set. As we have already seen in the Japanese loanword devoicing example, the greater power of weighting can allow a set of constraints to capture an attested language that OT would fail to generate. It is thus likely that for any fleshed out theory of some domain, the HG constraint set will be different from the one assumed in OT. When HG and OT are operating with different constraint sets, there is no necessary relation of relative restrictiveness.

A related inference that one might draw is the one stated by Coetzee and Pater (2005: 114):

- (26) One of the most striking results of the typological research that has been conducted in Optimality Theory is that there seems to be very little counter-evidence for strict domination.

That is, the large body of successful linguistic analysis with the more restrictive ranked constraints stands as an argument for the premise that ranking suffices (a similar observation is made by Prince and Smolensky 1993/2004: 94). The flaw in this argument is that OT can capture any pattern that HG can, if we expand the constraint set appropriately (see also the discussion of Japanese loanword devoicing above).

A comparison of the two frameworks then must necessarily involve careful analysis of attested languages with both weighted and ranked constraints, and a comparison of the resultant typological predictions. This sort of research is just in its infancy. Jesney (2012, this volume) shows that HG can generate more of the attested typology of positional restrictions with positional markedness constraints than OT can, raising the possibility of avoiding unwelcome typological predictions of positional faithfulness. Potts *et al.* (2010) provide a similar comparison of an HG analysis of Lango vowel harmony with an OT-LC one, again showing that the more restricted constraint set in HG generates a tighter typology than OT with a less restricted constraint set.

Scalar constraints are known to undergenerate in OT (Prince and Smolensky 1993/2004, McCarthy 2003). In Pater (2012), I show that HG allows them to capture the basic type of attested pattern that makes them incompatible with OT. Here I use that case to provide an explicit example of how OT and HG operating with different constraint sets produce partially overlapping typological predictions. I also discuss some of the issues in making scalar constraints accountable for existing data patterns. The investigation of scalar constraints in HG began with Flemming (2001), who uses HG to allow for real-valued scales, a distinct, but

closely related application of weighted constraints – see more recently Cho (2011), McAllister Byun (2011) and Ryan (2011) on weighted scalar constraints.

We start with a review of the example from Pater (2012): the differences between the sets of languages generated by HG and OT with a constraint on the sonority of syllable nuclei. McCarthy (2003) suggests the following restatement of Prince and Smolensky's (1993/2004) H-NUC, which I rename \*C-NUC in Pater (2012), to make it clear that the two constraints have different effects.<sup>5</sup>

(27) \*C-Nuc

Assign a violation mark to a nucleus for each degree of sonority separating it from [a]

Because \*C-Nuc assigns multiple violations to a single structure (that is, it is gradient, in McCarthy's 2003 sense), we can get asymmetric trade-offs between just it and another constraint. Assuming a simple sonority scale in which sonorant consonantal nuclei are one step away from [a] and obstruents are two, we get the \*C-Nuc violations for [tN] and [tS] in (28), where capitalization indicates nuclear status. The comparisons in (28) are with candidates with epenthesis of the vowel [a], which avoids a \*C-Nuc violation. These violate Dep, so the faithful candidates are preferred by a margin of +1.

(28) Comparative vectors for consonantal nuclei and epenthesis

W ~ L	*C-NUC	DEP
a. [tN] ~ [tan]	-1	+1
b. [tS] ~ [tas]	-2	+1

In OT, there are only two possibilities: either \*C-Nuc dominates Dep, and both epenthetic candidates in (28) win, or Dep dominates \*C-Nuc, and neither does. That is, OT cannot generate a cut-off in this scale. It is this problem that leads Prince and Smolensky (1993/2004: sec. 8.1) to replace the scalar H-Nuc with constraints in a fixed ranking, and McCarthy (2003) to propose a general prohibition against scalar constraints in OT.

HG does not suffer from this problem, as illustrated by the weighted comparative vectors in (29). The final column shows the weighted sum of violation differences, or margin of separation. Because these numbers are positive, the Winners are correctly preferred. With these weights, a penalty of -2 on \*C-Nuc, as would be incurred by the sub-optimal \*[tS], is sufficient to force epenthesis, but the penalty of -1, as assessed for [tN], is not. English is a close-by example of a language that imposes this sort of restriction, though a full account would need to deal with the fact that only syllabic [ɹ] is permitted in stressed syllables.

(29) Weighted comparative vectors for a language with only sonorant nuclei

W ~ L	*C-NUC	DEP	
	2	3	
a. [tN] ~ [tan]	-1	+1	+1
b. [tas] ~ [tS]	+2	-1	+1

While scalar constraints in HG do not have the basic problem that they do in OT, there are more subtle issues that remain to be dealt with. De Lacy (2004) discusses a pattern he calls

“conflation”, where some sonority differences fail to play a role in a given language. He provides as one example stress placement in Nganasan. In words with three light syllables, stress generally falls on the penultimate syllable. However, when the penult is of low sonority, containing one of the vowels [i y u ə i], and the antepenult is of higher sonority, containing mid [e o] or low [a], stress falls on the higher sonority antepenult (30a.). Conflation is observed in that not all sonority differences upset the basic stress pattern in this way. Antepenult low vowels fail to attract stress away from penult mid ones (30b.), and antepenult high vowels fail to attract stress from penult central ones (30c.). De Lacy shows that in other languages, there are sonority sensitive stress patterns that distinguish higher sonority low vowels from mid ones, and high from central.

- (30) a. [cétəmti] ‘fourth’; [kóntuʃa] ‘carry’; [négyʃa] ‘tease’; [ániʔə] ‘large’; [hʰásirə] ‘fishing rod’  
 b. [bacébsa] ‘breathing’; [kacéməʔ] ‘examine’; [lʷamóbtuʔ] ‘spill, splash’  
 c. [cintáʃi] ‘stroke’; [kunsíni] ‘inside’; [hytódə] ‘torso’

If scalar constraints assign violations according to universal linear scales, this sort of pattern cannot be accounted for. We will focus just on the low/mid/high distinctions, and assume the constraints in (31) and (32):

- (31) Stress-to-Sonority-Linear (Stress-Son-Lin)

Assign a violation to the head of a foot for each degree of sonority separating it from [a]  
 (low vowel = 0, mid vowel = -1, high vowel = -2)

- (32) Penult

Assign a violation for antepenultimate stress

The problem is that the difference between the penalties assigned to stressed high vowels (-2) vs. stressed mid ones (-1) is the same as the difference between those assigned to stressed mid vowels (-1) vs. low ones (0). This leads to the inconsistent weighting conditions illustrated in (33): antepenultimate stress in [négyʃa] requires STRESS-SON-LIN to outweigh Penult, but penultimate stress in [kacéməʔ] requires the reverse.

- (33) Inconsistent comparative vectors with a linear scalar constraint

W ~ L	STRESS-SON-LIN	PENULT
a. [négyʃa] ~ [negýʃa]	+1	-1
b. [kacéməʔ] ~ [káceməʔ]	-1	+1

One solution would be to allow language-specific conflation in the violation scales. For example, in Nganasan mid vowels would be conflated with low vowels in not violating STRESS-SON-LIN. Another solution is to adopt nonlinear violation scales. With the revised STRESS-SON-NONLIN in (34), the distance between mid and high is 2, while the distance between low and mid remains 1.

- (34) Stress-to-Sonority-Nonlinear (STRESS-SON-NONLIN)

Assign a violation to the head of a foot for each degree of sonority separating it from [a]  
 (low vowel = 0, mid vowel = -1, high vowel = -3)

This revision allows us to find a correct set of weights, as shown in (35).

## (35) Weighted comparative vectors with a nonlinear scalar constraint

W ~ L	STRESS-SON-NONLIN	PENULT	
	2	3	
a. [négyʃa] ~ [negýʃa]	+2	-1	+1
b. [kacéməʔ] ~ [káceməʔ]	-1	+1	+1

De Lacy (2004) points out that some patterns of conflation are a problem for fixed rankings of constraints in OT, and argues that constraints in a stringency relation can get the full range of patterns. The constraints required in OT for the patterns we have been analyzing with the scalar Stress-Son constraints are given in (36).

## (36) OT Stress-Sonority constraints

a. \*Stress-[i]

Assign a violation to a head of a foot that is a high vowel

b. \*Stress-[i,e]

Assign a violation to a head of a foot that is a high or mid vowel

We can thus compare OT with the set of constraints in (36) to HG with the constraint in (34), each interacting with the Penult constraint. The first row of the table in (37) shows three inputs, which illustrate the three possible distributions of penultimate and antepenultimate vowel sonority in which the antepenult has greater sonority than the penult. The first two columns differ in whether the difference is between mid and high (/teniti/) or low and mid (/taneti/); the third column pits low against high (/taniti/). The rows beneath show the output stress patterns generated by HG and OT with different constraint sets. Asterisks indicate the instances of antepenultimate stress, where the demands of a stress-sonority constraint prevail over the penult positional preference. Checkmarks indicate which theories generate each pattern. Our present comparison is between OT and HG-1.

## (37) Typological predictions of OT and HG with different constraint sets

	/teniti/	/taneti/	/taniti/	OT	HG-1	HG-2
a.	téniti*	táneti*	tániti*	✓	✓	✓
b.	teníti	taníti	taníti	✓	✓	✓
c.	téniti*	tanéti	táneti*	✓	✓	✓
d.	teníti	táneti*	tániti*	✓		✓
e.	teníti	taníti	tániti*		✓	✓

The pattern in (37a.) is one in which sonority fully determines the outcome, and the one in (37b.) is the converse case where the positional preference is always respected. The next row, (37c.) is the Nganasan conflation pattern, where low and mid vowels attract stress from high ones, but in which the low-mid difference is ignored.

The last two rows of (37) illustrate OT-HG differences. As the checkmarks indicate, OT is more restrictive in one way, and HG in another. The difference illustrated in (37e.) arises from the fundamental difference between ranked and weighted constraints: weighted constraints can model a “sufficient reward” threshold, in which a general preference is overridden only to gain a sufficient benefit on another dimension. Here that benefit is stressing the best type of vowel, low [a], instead of the worst one, high [i]. A gain from worst to intermediate –high [i] to

mid [e] – or intermediate to best – mid [e] to low [a] – is insufficient to compensate for placing stress on the dispreferred antepenultimate position. De Lacy's (2004, 2006) typological survey appears to include no vowel quality-based stress pattern with a sufficient reward threshold. Further research is required to determine whether that gap is accidental, as predicted by HG, or is a reflection of a general restriction on constraint interaction, as predicted by OT. For sonority-stress interactions, it seems that the existing set of typological data is too sparse to make this determination.

The row in (37d.) shows that HG fails to generate a pattern in which conflation is between mid and high vowels, rather than between low and mid. This is an example of a prediction that derives from nonlinear scalar constraints. When there is a greater difference in the number of violations on one step of the scale than another (e.g. with STRESS-SON-NONLIN, low-mid = 1, mid-high = 2), if the smaller difference favors an optimum, so will the larger difference (all else being equal). Here, it is impossible to make [táneti] beat [tanéti] while at same time making [teníti] beat [téniti], as the inconsistent comparative vectors in (38) show.

(38) Inconsistent comparative vectors for mid-high conflation

W ~ L	STRESS-SON-NONLIN	PENULT
a. [táneti] ~ [tanéti]	+1	-1
b. [teníti] ~ [téniti]	-2	+1

For this particular example, the difference counts in favor of OT: de Lacy (2006) points to Gujarati as an example of a language that conflates the sonority scale in this way in arbitrating stress placement. The overall resulting OT prediction, that there are no implicational relationships of this type in conflation, remains to be critically examined. *A priori*, it seems plausible that some adjacent points on scales, sonority and otherwise, are more easily conflated than others, and nonlinearities in the scales seem a like natural way to express this.

The column labeled “HG-2” in (37) shows the result if we include two nonlinear Stress-to-Sonority constraints, one that has a greater difference between mid and high vowels, and one that has a greater difference between low and mid. These are labeled Stress-to-Sonority-13 and Stress-to-Sonority-23 in (39) respectively. With these constraints, HG gets the full set of patterns.

(39) a. Stress-to-Sonority-13

Assign a violation to the head of a foot for each degree of sonority separating it from [a] (low vowel = 0, mid vowel = -1, high vowel = -3)

b. Stress-to-Sonority-23

Assign a violation to the head of a foot for each degree of sonority separating it from [a] (low vowel = 0, mid vowel = -2, high vowel = -3)

While this may appear to be a brute force solution, mid vowels are independently known to vary in their featural affiliation with low and high vowels, hence their usual designation as [-high, -low], with features that allow them to be classed with either.

To further paint what is likely to be the general picture of partially overlapping typological predictions between OT and HG with different constraint sets, we can add competing constraints that apply to another scalar dimension: syllable weight. Languages typically make only a two-way distinction between heavy and light syllables, but some languages make a

three-way distinction. For example, in the variety of Hindi described by Kelkar (1968) and analysed by Hayes (1995) and Prince and Smolensky (1993/2004), syllables with long vowels or codas count as heavy, and syllables with both of these count as superheavy. For HG, the typology calculation used the scalar Weight-to-Stress constraint in (40), interacting with the two Stress-to-Sonority constraints in (39). The nonlinear scale in (40) makes conflation between heavy and light entail conflation between heavy and superheavy.

(40) Weight-to-Stress

Assign a violation to an unstressed heavy syllable (–2 if heavy, –3 if superheavy)

For OT, the weight-to-stress constraints in the typology calculation are the ones in (41), which interact with the stress-to-sonority constraints in (36).

(41) a. Heavy-to-Stress

Assign a violation to an unstressed heavy or superheavy syllable.

b. Superheavy-to-Stress

Assign a violation to an unstressed superheavy syllable.

The results are illustrated in (42). For every input, the syllable on the left is heavier than the syllable on the right, while the one on the right is higher in sonority. The first six languages are generated both by OT and HG with their respective constraint sets. The first language is one in which the sonority-sensitive constraints fully determine the outcome, and the last of the six is one in which the weight-sensitive constraints are fully obeyed. In between those two are the 4 languages that display intermediate outcomes that both theories can generate, while in the last 19 languages are 9 mixed patterns that only OT can produce, and the 10 that only HG gets.

## (42) OT and HG with different constraint sets

ti:nma:	ti:nma	ti:me	te:nma:	ti:nme	te:nma	te:ma	ti:ma	ti:nme:	OT	HG
ti:nmá:	ti:nmá	ti:mé	te:nmá:	ti:nmé	te:nmá	te:má	ti:má	ti:nmé:	✓	✓
ti:nmá:	ti:nmá	ti:mé	te:nmá:	ti:nmé	té:nma	té:ma	ti:má	ti:nmé:	✓	✓
ti:nmá:	ti:nmá	tí:me	te:nmá:	tí:nme	te:nmá	te:má	ti:má	ti:nmé:	✓	✓
ti:nmá:	tí:nma	tí:me	te:nmá:	tí:nme	té:nma	té:ma	tí:ma	tí:nme:	✓	✓
ti:nmá:	tí:nma	tí:me	té:nma:	tí:nme	té:nma	té:ma	tí:ma	ti:nmé:	✓	✓
tí:nma:	tí:nma	tí:me	té:nma:	tí:nme	té:nma	té:ma	tí:ma	tí:nme:	✓	✓
ti:nmá:	ti:nmá	ti:mé	te:nmá:	tí:nme	te:nmá	te:má	ti:má	tí:nme:	✓	
ti:nmá:	ti:nmá	ti:mé	té:nma:	ti:nmé	té:nma	te:má	ti:má	ti:nmé:	✓	
ti:nmá:	ti:nmá	ti:mé	té:nma:	ti:nmé	té:nma	té:ma	ti:má	ti:nmé:	✓	
ti:nmá:	ti:nmá	tí:me	te:nmá:	tí:nme	te:nmá	te:má	ti:má	tí:nme:	✓	
ti:nmá:	tí:nma	tí:me	te:nmá:	tí:nme	té:nma	té:ma	tí:ma	ti:nmé:	✓	
ti:nmá:	tí:nma	tí:me	té:nma:	tí:nme	té:nma	té:ma	tí:ma	ti:nmé:	✓	
tí:nma:	tí:nma	ti:mé	té:nma:	tí:nme	té:nma	te:má	ti:má	tí:nme:	✓	
tí:nma:	tí:nma	ti:mé	té:nma:	tí:nme	té:nma	té:ma	ti:má	tí:nme:	✓	
tí:nma:	tí:nma	tí:me	té:nma:	tí:nme	té:nma	te:má	ti:má	tí:nme:	✓	
ti:nmá:	tí:nma	tí:me	té:nma:	tí:nme	té:nma	té:ma	tí:ma	tí:nme:		✓
ti:nmá:	tí:nma	tí:me	te:nmá:	tí:nme	té:nma	té:ma	ti:má	tí:nme:		✓
ti:nmá:	tí:nma	tí:me	té:nma:	tí:nme	té:nma	té:ma	ti:má	ti:nmé:		✓
ti:nmá:	tí:nma	tí:me	te:nmá:	tí:nme	té:nma	té:ma	ti:má	ti:nmé:		✓
ti:nmá:	ti:nmá	ti:mé	te:nmá:	tí:nme	té:nma	té:ma	ti:má	ti:nmé:		✓
ti:nmá:	ti:nmá	ti:mé	te:nmá:	tí:nme	té:nma	te:má	ti:má	ti:nmé:		✓
ti:nmá:	ti:nmá	ti:mé	te:nmá:	tí:nme	té:nma	te:má	ti:má	ti:nmé:		✓
ti:nmá:	ti:nmá	ti:mé	te:nmá:	tí:nme	te:nmá	te:má	ti:má	ti:nmé:		✓

This typological study illustrates the general way in which OT and HG with differing constraint set can generate partially overlapping predictions, and some of the particular differences between OT with constraints in a stringency relation, and HG with explicitly scalar constraints. Unfortunately, currently available typological data do not seem to choose one of these theories of constraints and their interaction over the others, since both sonority sensitivity and scalar weight sensitivity are rare. The choice between HG and OT will likely thus have to be made on the basis of other data.

#### 4 Unbounded trade-offs and locality

Legendre, Sorace and Smolensky (2006) provide an example of an unattested linguistic system produced by HG but not OT. The example involves what can be referred to as an unbounded trade-off: satisfaction of one constraint can require a potentially unbounded number of violations of another. The constraints at issue are ones that determine stress placement. The first requires that a particular kind of syllable – a heavy one – be stressed (Weight-to-Stress; Prince, 1990). Languages vary in which syllables fall into the heavy category; for the abstract example below, heavy syllables are ones that have codas. The other constraint penalizes a stressed syllable according to how far away it is from a word edge: it assigns a violation mark for each syllable that intervenes. I adopt the name MainStressRight from Legendre, Sorace and

Smolensky (2006) (cf. Edgemost in Prince and Smolensky 1993/2004, and Alignment in McCarthy and Prince 1993).

The table in (43), adapted from Legendre, Sorace and Smolensky (2006), compactly illustrates the unbounded trade-off. The syllable [ban] stands in for any heavy syllable and the coda-less syllable [ta] stands in for any nonheavy one. The variable  $\sigma_n$  is a string of a number  $n$  of nonheavy syllables. As in Legendre, Sorace and Smolensky (2006), we only consider candidates with a single stress per word. Stress on the final syllable violates Weight-to-Stress and satisfies MainStressRight. Stress on the initial syllable satisfies Weight-to-Stress and violates MainStressRight once for every syllable separating it from the right edge of the word. Stress on any of the syllables in  $\sigma_n$  would be harmonically bounded by final stress, as it would add at least one violation of MainStressRight without compensating improvement on Weight-to-Stress. Therefore, we need only consider the two candidates in (43).

(43) An unbounded trade-off

ban $\sigma_n$ ta	WEIGHT-TO-STRESS	MAINSTRESSRIGHT
a. ban. $\sigma_n$ .tá	-1	
b. bán. $\sigma_n$ .ta		-1- $n$

As there is no theoretical upper bound on the size of words, there is theoretically no upper bound on the number of MainStressRight violations that can be traded off against the single violation of Weight-to-Stress.

Legendre, Sorace and Smolensky (2006) point out that in OT there are only two possible languages given by the two rankings. The number of syllables intervening between a nonfinal stressed heavy syllable and the edge of the word is irrelevant; either stress will fall on the rightmost syllable or on the heavy syllable.

(44) Two languages in OT

Weight-to-Stress  $\gg$  MainStressRight  $\rightarrow$  bán. $\sigma_n$ .ta

MainStressRight  $\gg$  Weight-to-Stress  $\rightarrow$  ban. $\sigma_n$ .tá

They also note that HG produces a theoretically infinite set of languages with these constraints. With appropriate weights, stress can be limited to a “window” of any number of syllables at the right edge of the word (see also Prince 1993, 2007 on Goldsmith 1994). In HG, the number of intervening syllables is crucial, as shown by the fact that this number is included in the weighting conditions in (45). In these inequalities, the constraint names stand for their weights. Stress will fall on a nonfinal heavy syllable only if the weight of Weight-to-Stress is greater than the weight of MainStressRight times the number of syllables separating the heavy syllable from the right edge. For example, if Weight-to-Stress = 3.5 and MainStressRight = 1, then a heavy syllable will get stressed if it is followed by three light syllables, but not four. No known language has such a four-syllable window.

(45) An infinite typology in HG

Weight-to-Stress  $>$   $(n+1)$ \*MainStressRight  $\rightarrow$  bán. $\sigma_n$ .ta

$(n+1)$ \*MainStressRight  $>$  Weight-to-Stress  $\rightarrow$  ban. $\sigma_n$ .tá

Taken on its own, this case is not particularly persuasive as an argument against HG as a framework for typological study. One problem is that there are, in fact, attested three-syllable

windows (e.g., in Macedonian: Comrie, 1976; and Pirahã: Everett & Everett, 1984), and these cannot be generated by OT rankings of the set of constraints for stress in Prince and Smolensky (1993/2004) and McCarthy and Prince (1993) – see Hyde (2007) and Kager (to appear). One might take the ability of HG to account for them as a positive result and seek an explanation for the absence of the larger windows, perhaps in terms of the size of words that a learner would need to hear to acquire the pattern (cf. Hammond, 1991), or in terms of the relative difficulty of acquiring the weight ratios needed to represent the pattern (cf. Prince, 1993: 91; Prince, 2007a: 41).<sup>6</sup>

Another problem is that gradient Alignment constraints are controversial, even in OT. In counting the distance between two portions of the representation, these constraints assign violation scores in an unusual manner, both in comparison with other OT constraints, and with ones elsewhere in GL (Eisner 1998, Potts and Pullum 2002, Bíró 2003, and McCarthy 2003). For the other constraints discussed in this paper, which are typical of OT, each locus of violation incurs a bounded number of violation marks (see McCarthy 2003 on the formalization of “locus of violation”). For a stressed syllable evaluated by a MainStressRight, the number of violations depends on the distance from the edge of the word, and is thus unbounded. Not only are the gradient Alignment constraints formally unusual, but they also produce undesired typological predictions, in OT as well as HG. Based on these considerations, McCarthy (2003) proposes a revised theory of OT constraints in which gradient Alignment constraints are banned. Removing gradient Alignment constraints from OT has the effect of also removing a large class of potential problem cases for a version of the theory with weighted constraints.

Unbounded trade-offs, and divergences between the typological predictions of OT and HG, can also emerge from the interaction of constraints that only assign a single violation per locus. In some situations, the satisfaction of an output constraint can require a number of faithfulness violations with no theoretical upper bound. As an example, we can consider the interaction of NoCoda with the faithfulness constraint Linearity (McCarthy and Prince 1999), which assigns a violation mark for every pairwise reordering of the segmental string.

(46) Linearity

If segment *x* precedes segment *y* in the input, *x* precedes *y* in the output

In (47), word-internal syllable boundaries are again indicated with periods, and NoCoda violations again occur when a syllable ends in a consonant (clusters are assumed to be split between syllables). Here we see that satisfaction of NoCoda can require two reorderings of the segmental string, as (47d).

(47) Asymmetric trade-off between NoCoda and Linearity

/apekto/	NoCODA	LINEARITY
a. [a.pək.to]	–1	
b. [pa.ək.to]	–1	–1
c. [ap.ke.to]	–1	–1
d. [pa.ke.to]		–2

It is also possible to create strings in which only one violation of Linearity would be needed to satisfy NoCoda (e.g., /ekto/, [ke.to]), as well as ones in which any higher number is needed

(e.g., /idapekto/ requires three violations, as in [di.pa.ke.to]). Appropriate weightings of the constraints can create systems in which NoCoda is satisfied at the cost of  $n$  violations of Linearity, but not  $n + 1$  violations, where  $n$  is any nonnegative integer. OT only produces two systems: one in which NoCoda is satisfied at any cost in terms of Linearity violations, and one in which even a single Linearity violation is worse than a violation of NoCoda.

McCarthy (2007a) in fact provides this example as a case in which the standard version of OT produces the wrong result. Although languages do employ local pairwise reorderings of segments to satisfy output constraints like NoCoda (see Hume, 2001 for a survey), none use a double reordering of the type illustrated in the final candidate in (47), which would be optimal under the ranking NoCoda  $\gg$  Linearity (as well as with weights respecting the condition NoCoda  $>$  2\*Linearity).

McCarthy (2007a) shows that the correct typology is obtained in the alternative version of OT that Prince and Smolensky (1993/2004) call Harmonic Serialism (HS) – see McCarthy (this volume) for an overview of other typological advantages of HS. In this theory, Gen is limited to a single application of an operation; here it can produce a single pairwise reordering, but not two at once. Multiple applications of an operation can occur serially, if each one results in an improvement in harmony. The tableau in (48) shows the first candidate set that would be produced in an HS derivation, which lacks the candidate with a double reordering. The faithful candidate (48a.) harmonically bounds the others, so it would always be picked as optimal regardless of the constraint ranking. In HS a derivation terminates when the faithful candidate is chosen; (48) is thus both the first and last step.

(48) No double reordering in Harmonic Serialism

/apekto/	NoCODA	LINEARITY
a. $\rightarrow$ [a.pək.to]	-1	
b. [pa.ək.to]	-1	-1
c. [ap.ək.to]	-1	-1

McCarthy's solution extends to a weighted constraint version of the theory, as [a.pək.to] is equally guaranteed to win in (48) with any set of positive weights. It also eliminates the difference between OT and HG typology mentioned beneath (47), since in a serial version of either OT or HG, only single reorderings (e.g. /ekpo/, [ke.po]) can be used to satisfy NoCoda.

We now return the example that Prince and Smolensky (1997) point to as distinguishing ranked from weighted constraints: the interaction of NoCoda and Parse. In section 2.3, we saw that the number of potential codas did not in fact lead to a difference between HG and OT: in either theory, all are deleted, or all are retained. A difference can emerge, though, from the second part of their example: the number of potential consonants in a single coda. If we adopt a version of NoCoda that assigns only single violation for the entire coda, we can create a system in HG in which NoCoda is satisfied at the cost of  $n$  violations of Parse, but not  $n + 1$ .

As the pair of tableau in (49) and (50) shows, the result of the interaction of these constraints in parallel HG can be quite bizarre: a language that has codas with two or more consonants (e.g.  $\rightarrow$  [(apt)] in (50)), but not one (e.g. \*[(ap)] in (49)). Further, this cut-off can be made at any point: languages with codas with no fewer consonants than three, or four, or five, or any other number can be modeled in this theory.

- (49) A single consonant fails to be parsed into a coda

	ap	NOCODA	PARSE	
		1.5	1	
a. →	(a)p		-1	-1
b.	(ap)	-1		-1.5

- (50) Two consonants are parsed into a coda

	apt	NOCODA	PARSE	
		1.5	1	
a.	(a)pt		-2	-2
b.	(ap)t	-1	-1	-2.5
c. →	(apt)	-1		-1.5

If consonants are added one at a time to a syllable with an adjunction operation, as in Pater (2012), then this pattern is impossible to recreate in a serial version of HG.<sup>7</sup> With our two constraints, the first application of adjunction will beat the fully faithful candidate if and only if NOCODA has a greater weight than PARSE-SEG. The tableau in (51) illustrates the second step of the derivation for the UR /apt/ with the same weights as in (50). Here we already have the nucleus syllabified through the prior application of nuclear projection. Importantly, the candidate set includes only the faithful candidate and the single adjunction, and not the fully syllabified candidate (apt) that was optimal in the parallel HG tableau for /apt/ in (50). Since that candidate is missing from the candidate set, the optimum is now [(a)pt]; we no longer get the strange pattern in which a coda is formed only to syllabify some minimum number of segments.<sup>8</sup>

- (51) Coda formation in serial HG

	(a)pt	NOCODA	PARSE	
		1.5	1	
a. →	(a)pt		-2	-2
b.	(ap)t	-1	-1	-2.5

This example illustrates a general consequence of the single change limitation on candidates imposed in a serial theory: that the set of possible constraint interactions, or trade-offs in the terminology of section 2, is restricted relative to a parallel model. In particular, trade-offs involving multiple instances of violation of a given faithfulness constraint will never occur in a serial version of HG or OT, insofar as the operations that create candidates incur at most one violation of each faithfulness constraint (as in McCarthy 2007b). Thus, scenarios in which a markedness constraint is satisfied at the cost of  $n$  violations of a faithfulness constraint, but not  $n + 1$ , which can be created in a parallel version of HG, are impossible in a serial version. The typological benefit accrued by serialism in the \*CODA/PARSE example is likely typical of such cases. For example, a parallel version of HG with Agree constraints for assimilation can generate a pattern in which one of a sequence of two disagreeing segments will assimilate, but

in which assimilation is blocked if some larger number of segments would have to be changed to achieve agreement across the whole string (Pater, Bhatt and Potts 2007, Bane and Riggle 2012). This sort of pattern does not arise in the serial HG approach to assimilation developed by Mullin (2012).

Both gradient Alignment and the parallel evaluation of multiple instances of faithfulness constraint violation can be characterized as producing unwanted global effects in the standard version of OT. We thus have unwanted globality introduced by particular types of constraint, and by particular assumptions about how candidates are generated and evaluated. In both cases, it seems likely that refinements to the theory that impose the desired locality restrictions in OT will result in the elimination of the most serious problems faced by weighted constraint versions of OT-style UG. The biggest challenge faced by HG appears to be in stress typology. Although as discussed above, the particular case of an HG/OT difference discussed by Legendre, Sorace and Smolensky (2006) does not incontrovertibly favor OT, the full predicted typology generated by parallel HG with the alignment constraints of McCarthy and Prince (1993) does not look to be particularly promising for the success of this theory (see further Potts *et al.* 2010 and Bane and Riggle 2012). If the right solution to the other locality problems discussed in section 4 is to adopt HS, then we run into the problem that gradient Alignment seems to be necessary in a standard serial theory (Pruitt 2012).<sup>9</sup>

## 5 Probabilistic HG

It appears to be the standard assumption in current phonological theory that a realistic model of phonology must be able to cope with variation in outcomes across instances of production (and/or perception and/or experimental judgment) – see Coetzee and Pater (2011) for a recent overview of the data that motivate this view, and of OT and HG models of grammar that have stochastic output. There are two versions of probabilistic HG currently being pursued: Maximum Entropy Grammar (MaxEnt; Goldwater and Johnson 2003, Wilson 2006, Jäger and Rosenbach 2006, Jäger 2007, Hayes, Zuraw, Siptár and Londe 2009, Staubs and Pater this volume) and Noisy HG (see Boersma and Pater this volume for a description and references). These theories have gained attention primarily because of the existence of their associated learning algorithms. In this section I discuss how the behavior of their grammatical components relates to the general properties of the categorical version of HG explored in this paper.

In the categorical version of HG, the candidate with the highest Harmony, or weighted sum of constraint violations, is chosen as the optimum. This choice does not change across instances of evaluation. In Noisy HG, at each instance of evaluation the constraint weights are sampled from normal distributions around mean values, and the candidate with highest Harmony is chosen as optimal – because of the sampled weights, the optimum can vary across instances of evaluation. In MaxEnt the probability of a candidate is proportional to the exponential of its Harmony, and each time the grammar is used a candidate is sampled from this distribution. In both of these probabilistic versions of HG, the probability of a candidate increases with its (mean) Harmony, so to make one candidate more probable than another, it must be given higher Harmony. Since the relative Harmony of candidates is always at issue, by understanding the patterns produced by categorical HG, we are also learning about the patterns that can be produced by the probabilistic theories.

If we limit a MaxEnt or Noisy HG grammar to positive weights, then the sets of single optima that the categorical version of HG can generate are the same sets of single candidates to

which the stochastic theories can give highest probability within their candidate sets, putting aside ties. Thus, we can do fairly direct mapping from the typological predictions of categorical HG discussed above. All and only the languages generated by categorical HG are generated by the probabilistic theories, where a “language” has a single optimal candidate in each tableau in the categorical case, and a single candidate with greatest probability in the probabilistic case.

MaxEnt gives non-zero probability to candidates that are harmonically bounded in categorical HG (Jäger and Rosenbach 2006, Jesney 2007), that is, to candidates that can never have higher Harmony than all of their competitors. For simply harmonically bounded forms – those that have a proper superset of the violation marks of another candidate – the resulting prediction is that they can only ever surface as a minority variant (see Goldrick and Daland 2009 for related discussion with respect to speech errors and a version of Noisy HG that permits negative weights). As usual, the prediction critically depends on the contents of the constraint set, which may make formulating a definitive test of this prediction of MaxEnt difficult. In Noisy HG, on the other hand, simply harmonically bounded forms always have zero probability, since they can never win in any single evaluation, no matter what the positive weights are.

Putting aside ties again for the moment, the candidates that are collectively bounded in HG also have zero probability in Noisy HG. The tableau in (52) demonstrates that MaxEnt can give collectively bounded candidates maximum probability within a candidate set. Candidates (52b) [dak.bad] and (52c) [dag.bat] are collectively bounded in HG in that there is no weighting that can make either of them singly optimal: with \*Coda-Voice > Ident-Voice, (52a) [dak.bat] wins, and with the reverse relationship, (52d) [dag.bat] does. The harmony of each candidate continues to be indicated as the value at the end of each row, and the MaxEnt probability assigned to each candidate is shown to its left. Under an equal weighting of the constraints, candidates (52b) and (52c) tie with (52a) and (52d), which translates into equal MaxEnt probability. This is as much probability as they can get: their probability decreases with respect to (52a) or (52d) as the weights are shifted in favor of \*Coda-Voice or Ident-Voice respectively – neither (52b) nor (52c) can ever gain higher probability than both (52a) and (52d).

(52) Collectively bounded candidates with maximum probability in MaxEnt

/dagbad/	*CODA-VOICE	IDENT-VOICE	
	2	2	
a. 0.25 dag.bat	-2		-4
b. 0.25 dak.bad	-1	-1	-4
c. 0.25 dag.bat	-1	-1	-4
d. 0.25 dak.bat		-2	-4

As Jäger and Rosenbach (2006) and Jesney (2007) point out, it may be to the advantage of MaxEnt that it can make candidates like (52b) and (52c) surface. For many processes, ‘local optionality’, as illustrated by devoicing of only one of two codas, is attested (see Kimper 2011 for a recent overview). However, as Jesney notes, the issue is complicated by the fact that there are alternative analyses of local optionality. In particular, Kimper (2011) develops a serial account in HS, and as discussed in section 4, HS has independent benefits for HG typology (see Staubs and Pater this volume on learning in serial probabilistic HG).

As mentioned at the end of section 2.1, I abstract from ties in optimality in all of the discussion of categorical HG. In categorical HG, tableau (52) would be an instance of such a tie, and this one would be impossible if the constraints were in a totally ordered ranking as in standard OT. If we were committed to a categorical theory, then we would need to say something about the greater tie-generating capabilities of HG than standard OT. When we move to a more realistic stochastic version of the theory, ties are unremarkable. In MaxEnt they are simply interpreted as equal probability, as in (52). In Noisy HG, they are vanishingly improbable because they require precise weight values (so 52b and 52c would have vanishingly little probability in Noisy HG), and when they occur, can be simply resolved by random choice, as in Boersma and Pater (this volume).

## 6 Conclusions

In this paper, I have aimed to reopen the discussion of whether weighted constraints are suitable as a framework for phonological analysis and typological study, a discussion which may have seemed to have ended when Prince and Smolensky (1993/2004) rejected weighted constraints in favor of ranked ones. I have tried to show that the choice between OT and HG is far more difficult to make than one might initially assume, and that HG has considerable untapped potential in this domain. The asymmetric trade-off requirement imposes inherent restrictions on the types of cumulative interaction that HG can express – restrictions that are not shared by the alternative formalization of cumulativity in OT with local constraint conjunction. In addition, the greater expressive power of weighted than ranked constraints allows for new hypotheses about the contents of the universal constraints set. The study of OT and HG operating with different constraint sets is still a new area of research, and the above comparison of OT and HG accounts of scalar phenomena indicates some of the open paths for further work in this area. Finally, a full tally of the relative benefits of weighed and ranked constraints must also take into account the simultaneous choice of parallel *vs.* serial candidate evaluation, since as we saw in the last section, the locality restrictions imposed by serialism eliminate some of the differences between HG and OT.

Throughout this paper I have been discussing the standard OT view of UG, that it includes a universal set of constraints. Most learnability research in OT, starting with the work presented in Tesar and Smolensky (2000), takes this constraint set as given to learners. An alternative view is that UG consists of constraint schema, which are filled on the basis of learning data, and this alternative assumption is the basis of another active line of research with weighted constraints. An explicit demonstration of the viability of this sort of constraint induction is provided by Hayes and Wilson (2008), who work with a version of Maximum Entropy Grammar that determines a probability distribution over the space of possible words, and thus functions as a model of phonotactics (see also Moreton 2010 and Pater to appear on constraint induction with weighted constraints). Even though the premises and goals of this research are quite different from those of standard OT (see Pater and Moreton 2012 for extended discussion), there is much to be gained for the building of models of constraint induction from a better understanding of the patterns that a given set of constraints can and cannot generate. Such an understanding can be gained by further examining the predictions of sets of constraints, whose universality may be taken as a useful idealization. A choice between ranking and weighting also has to be made for induced constraints (see Adriaans and Kager 2010 on constraint induction with OT), though the arguments for one or the other become quite subtle if one is not aiming to derive typological generalizations directly from the constraint set.

## References

- Adriaans, Frans and Kager, René (2010) Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62: 311-321.
- Albright, Adam, Magri, Giorgio, and Michaels, Jennifer (2008) Modeling doubly marked lags with a split additive model. In Harvey Chan, Heather Jacob, and Enkeleida Kapia (eds.), *BUCLD 32: Proceedings of the 32nd annual Boston University conference on language development* 36–47. Somerville, MA: Cascadilla Press.
- Bakovic, Eric (2000) *Harmony, dominance, and control*. Doctoral dissertation, Rutgers University.
- Bane, Max, and Riggle, Jason (2012) The typological consequences of weighted constraints. In Nikki Adams, Adam Cooper, Fay Parrill, and Thomas Wier (eds.), *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, 45 (13-27). Chicago, IL: University of Chicago.
- Barnes, Jonathan (2006) *Strength and weakness at the interface: positional neutralization in phonetics and phonology*. Berlin and New York: Mouton de Gruyter.
- Bíró, Tamas (2003) Quadratic alignment constraints and finite state optimality theory. In Proceedings of the workshop on finite-state methods in natural language processing (119–126). Budapest, Hungary. [Available at: <http://roa.rutgers.edu/>].
- Blevins, Juliette (2004). *Evolutionary Phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.
- Boersma, Paul. 2009. Some correct error-driven versions of the Constraint Demotion algorithm. *Linguistic Inquiry* 40: 667-686.
- Boersma, Paul and Pater, Joe. This volume. Convergence properties of a gradual learning algorithm for Harmonic Grammar.
- Cho, Hyesun. 2011. The timing of phrase-initial tones in Seoul Korean: a weighted-constraint model. *Phonology* 28: 293-330.
- Coetzee, Andries and Pater, Joe (2011) The place of variation in phonological theory. In John Goldsmith, Jason Riggle, and Alan Yu (eds.), *The Handbook of Phonological Theory* (2nd edn.) 401-431. Malden, MA: Blackwell.
- Comrie, Bernard (1976) Irregular stress in Polish and Macedonian. *International Review of Slavic Linguistics* 1: 227–240.
- Eisner, Jason (1998) FootForm decomposed: Using primitive constraints in OT. In B. Bruening (Ed.), *Proceedings of SCIL VIII, number 31 in MIT Working Papers in Linguistics* (115–143). Cambridge, MA: Department of Linguistics, MIT.
- Elfner, Emily (This volume) Stress-Epenthesis Interactions in Harmonic Serialism.
- Everett, D., and Everett, K. (1984) On the relevance of syllable onsets to stress placement. *Linguistic Inquiry* 15: 705–711.
- Farris-Trimble, Ashley (2008) Cumulative faithfulness effects: Opaque or transparent? In Ashley W. Farris-Trimble and Daniel A. Dinnsen (eds.) *IUWPL6: Phonological Opacity Effects in Optimality Theory* 119-145. Bloomington, IN: IULC Publications.

- Flemming, Edward (2001) Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7-44.
- Goldrick, Matthew and Daland, Robert (2009) Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology* 26: 147–185.
- Goldsmith, John (1990) *Autosegmental and metrical phonology*. Oxford and Cambridge, MA: Blackwell.
- Goldsmith, John (1993a). Harmonic phonology. In John Goldsmith (ed.) *The last phonological rule* (21–60). Chicago: University of Chicago Press.
- Goldsmith, J. (1993b). *The last phonological rule: Reflections on constraints and derivations*. Chicago: University of Chicago Press.
- Goldsmith, John (1994) A dynamic computational theory of accent systems. In Jennifer Cole and Charles Kisseberth (eds.), *Perspectives in phonology* (1–28). Stanford, CA: CSLI.
- Goldwater, Sharon and Johnson, Mark (2003) Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson and Östen Dahl (eds.), *Proceedings of the Stockholm Workshop on “Variation within Optimality Theory”* 111–120. Stockholm: Stockholm University.
- Guy, Gregory R. (1997). Violable is variable: Optimality Theory and linguistic variation. *Language Variation and Change* 9: 333–347.
- Hale, Kenneth (1973) Deep-surface canonical disparities in relation to analysis and change: An Australian example. In Thomas Sebeok (Ed.), *Current trends in linguistics, volume 9: Diachronic, areal and typological linguistics* 401–458. The Hague, The Netherlands: Mouton.
- Hammond, Michael (1991) Parameters of metrical theory and learnability. In Iggy Roca (ed.) *Logical issues in language acquisition* (47–62). Dordrecht, The Netherlands: Foris.
- Hayes, Bruce (1995) *Metrical stress theory: Principles and case studies*. Chicago: The University of Chicago Press.
- Hayes, Bruce, Robert Kirchner and Donca Steriade (eds.) (2004). *Phonetically based phonology*. Cambridge: Cambridge University Press.
- Hayes, Bruce, Tesar, Bruce and Zuraw, Kie (2003) *OTSoft 2.1*. Software package developed at UCLA.
- Hayes, Bruce and Wilson, Colin (2008) A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379–440.
- Hayes, Bruce Kie Zuraw, Péter Siptár, and Zsuzsa Londe (2009) Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language* 85: 822-863.
- Hume, Elizabeth (2001) Metathesis: Formal and functional considerations. In Elizabeth Hume, Norval Smith and Jeroen van de Weijer (eds.), *Surface syllable structure and segment sequencing* (1–25). Leiden, The Netherlands: HIL.
- Hyde, Brett (2007) An alignment approach to stress windows. Colloquium presentation, Rutgers University. [Available at: <http://www.artsci.wustl.edu/bhyde/windowshandout.pdf>.]
- Ito, Junko (1986) *Syllable theory in prosodic phonology*. Doctoral dissertation, University of Massachusetts, Amherst. Published 1988, New York: Garland.

- Ito, Junko and Mester, Armin (1986) The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17, 49–73.
- Ito, Junko and Mester, Armin (1995) Japanese phonology. In John Goldsmith (ed.), *Handbook of phonological theory* (817–838). Cambridge, MA: Blackwell.
- Ito, Junko and Mester, Armin (2003) *Japanese morphophonemics: Markedness and word structure*. MIT Press Linguistic Inquiry Monograph Series 41. Cambridge, MA: MIT Press.
- Jäger, Gerhard (2007) Maximum entropy models and Stochastic Optimality Theory. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Chris Manning (eds.), *Architectures, rules, and preferences: A Festschrift for Joan Bresnan* 467–479. Stanford, CA: CSLI.
- Jäger, Gerhard and Rosenbach, Anne (2006) The winner takes it all—almost. Cumulativity in grammatical variation. *Linguistics* 44: 937–971.
- Jesney, Karen (2012) Licensing in multiple contexts: an argument for Harmonic Grammar. In Nikki Adams, Adam Cooper, Fay Parrill, and Thomas Wier (eds.), *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 45 (287–301). Chicago, IL: University of Chicago.
- Jesney, Karen (This volume) Positional constraints in Optimality Theory and Harmonic Grammar.
- Jesney, Karen, and Anne-Michelle Tessier (2011) Biases in Harmonic Grammar: The road to restrictive learning. *Natural Language and Linguistic Theory* 29: 251–290.
- Kager, René (To appear) Stress in Windows: Language Typology and Factorial Typology. In *Lingua*.
- Kawahara, Shigeto (2011) Japanese loanword devoicing revisited: a rating study. *Natural Language and Linguistic Theory* 29: 705–723.
- Kelkar, Ashok R. (1968) *Studies in Hindi-Urdu. Vol. 1: Introduction and word phonology*. Poona: Deccan College.
- Kimper, Wendell A. (2011) Locality and globality in phonological variation. *Natural Language and Linguistic Theory* 29: 423–465.
- de Lacy, Paul (2004) Markedness conflation in Optimality Theory. *Phonology* 21: 145–199.
- de Lacy, Paul (2006) *Markedness*. Cambridge University Press.
- Legendre, Géraldine, Miyata, Yoshiro, and Smolensky, Paul (1990) Can connectionism contribute to syntax?: Harmonic grammar, with an application. Report CU-CS-485-90. Computer Science Department, University of Colorado at Boulder. [Available at: [www.cs.colorado.edu/departement/publications/reports/docs/CU-CS-485-90.pdf](http://www.cs.colorado.edu/departement/publications/reports/docs/CU-CS-485-90.pdf).]
- Legendre, Géraldine, Sorace, Antonella and Smolensky, Paul (2006) The Optimality Theory-Harmonic Grammar connection. In Paul Smolensky and Géraldine Legendre (eds.), *The harmonic mind: From neural computation to optimality theoretic grammar, vol. 2: Linguistic and philosophical implications* 339–402. Cambridge, MA: MIT Press.
- Lombardi, Linda (1991) *Laryngeal features and laryngeal neutralization*. Doctoral dissertation, University of Massachusetts, Amherst.
- Lubowicz, Anna (2002) Derived environment effects in Optimality Theory. *Lingua* 112: 243–280.

- Lubowicz, Anna (2005) Locality of conjunction. *WCCFL* 24. 254–262.
- Martin, Andrew (2007) *The evolving lexicon*. Doctoral dissertation, UCLA.
- McAllister Byun, Tara (2011) A gestural account of a child-specific neutralisation in strong position. *Phonology* 28: 371–412.
- McCarthy, John J. (1999). Sympathy and phonological opacity. *Phonology* 16: 331–399.
- McCarthy, John J. (2003) OT constraints are categorical. *Phonology* 20: 75–138.
- McCarthy, John J. (2007a) Restraint of analysis. In Sylvia Blaho, Patrick Bye and Martin Kramer (eds.), *Freedom of analysis* (203–231). Berlin: Mouton de Gruyter.
- McCarthy, John J. (2007b) *Hidden generalizations: phonological opacity in Optimality Theory*. London: Equinox.
- McCarthy, John J. (2009) Studying Gen. *Journal of the Phonetic Society of Japan* 13: 3–12. [Available at: [http://works.bepress.com/john\\_j\\_mccarthy/3](http://works.bepress.com/john_j_mccarthy/3)]
- McCarthy, John J. (This volume). The theory and practice of Harmonic Serialism.
- McCarthy, John J. and Prince, Alan (1993) Generalized Alignment. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology* 79–153) Dordrecht, The Netherlands: Kluwer. Excerpts appear in John Goldsmith ed. (1999), *Essential readings in phonology* (102–136). Oxford, England: Blackwell.
- McCarthy, John J. Prince, Alan (1999) Faithfulness and identity in Prosodic Morphology. In René Kager, Harry van der Hulst and Wim Zonneveld (eds.) *The prosody–morphology interface* (218–309). Cambridge, England: Cambridge University Press.
- Moreton, Elliott (2010) Constraint induction and simplicity bias in phonological learning. Paper presented at the Workshop on Grammar Induction, Cornell University [<http://www.unc.edu/~moreton/Papers/MoretonCornellWkshp2010.pdf>]
- Moreton, Elliott and Paul Smolensky (2002) Typological consequences of local constraint conjunction. *WCCFL* 21. 306–319.
- Mullin, Kevin (2011) Strength in Harmony Systems: Trigger and Directional Asymmetries. Ms., University of Massachusetts Amherst. [Available at <http://people.umass.edu/kmullin/Mullin2011StrengthHarmonySystems.pdf>]
- Nishimura, Kohei (2003) Lyman’s Law in loanwords. Master’s thesis, Nagoya University.
- Nishimura, Kohei (2006) Lyman’s Law in loanwords. *Phonological Studies* [Onin Kenkyuu], 9, 83–90.
- Padgett, Jaye (2002). Constraint conjunction versus grounded constraint sub-hierarchies in Optimality Theory. Ms, University of California, Santa Cruz. [Available as ROA-530 from the Rutgers Optimality Archive.]
- Pater, Joe (1999). Austronesian nasal substitution and other NC effects. In Kager, René, Harry van der Hulst and Wim Zonneveld (eds.) (1999). *The prosody-morphology interface*. Cambridge: Cambridge University Press. 310–343.
- Pater, Joe (2009a) Weighted Constraints in Generative Linguistics. *Cognitive Science* 33: 999–1035.
- Pater, Joe (2009b) Review of “The harmonic mind: From neural computation to optimality theoretic grammar”. *Phonology* 26: 217–226.

- Pater, Joe (2010) Non-vacuous specific with general gang effects. Ms, University of Massachusetts, Amherst. [Available at <http://roa.rutgers.edu/>.]
- Pater, Joe (2012) Serial harmonic grammar and Berber syllabification. In Toni Borowsky, Shigeto Kawahara, Takahito Shinya, & Mariko Sugahara (eds.), *Prosody matters: Essays in honor of Elisabeth O. Selkirk* (43-72), London: Equinox Press. [Available at <http://roa.rutgers.edu/>.]
- Pater, Joe (To appear) Canadian raising with language-specific weighted constraints. In *Language*.
- Pater, Joe, Rajesh Bhatt, and Christopher Potts (2007) Linguistic optimization. Ms, University of Massachusetts Amherst. [Available at <http://roa.rutgers.edu/>.]
- Pater, Joe and Elliott Moreton. 2012. Structurally biased phonology: Complexity in learning and typology. *EFL Journal*: 1-44. [Available at <http://people.umass.edu/pater/pater-moreton-2012-offprint.pdf>]
- Potts, Christopher, Pater, Joe, Jesney, Karen, Bhatt, Rajesh, and Becker, Michael (2010) Harmonic Grammar with Linear Programming: From linear systems to linguistic typology. *Phonology* 27: 77-117.
- Potts, Christopher and Pullum, Geoffrey K. (2002) Model theory and the content of OT constraints. *Phonology* 19: 361-393.
- Prince, Alan (1993) *In defense of the number i: Anatomy of a linear dynamical model of linguistic generalizations*. RuCCS Technical Report 1. New Brunswick, NJ: Rutgers University.
- Prince, Alan (2002) Entailed ranking arguments. Ms, Rutgers University. [Available at <http://roa.rutgers.edu/>.]
- Prince, Alan (2003) Anything goes. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata and Shin-ichi Tanaka (eds.) *A New Century of Phonology and Phonological Theory. A Festschrift for Professor Shosuke Haraguchi on the Occasion of His Sixtieth Birthday* 66–90. Tokyo: Kaitakusha. [Available at: <http://roa.rutgers.edu/>.]
- Prince, Alan (2007) In pursuit of theory. In Paul de Lacy (ed.), *Cambridge handbook of phonology* (33–60). Cambridge, England: Cambridge University Press.
- Prince, Alan and Paul Smolensky (1997) Optimality: from neural networks to Universal Grammar. *Science* 275: 1604–1610.
- Prince, Alan and Paul Smolensky (2004). *Optimality Theory: constraint interaction in generative grammar*. Malden, Mass. and Oxford: Blackwell. [Revision of 1993 technical report, Rutgers University Center for Cognitive Science. [Available at: <http://roa.rutgers.edu/>.]
- Pruitt, Kathryn (2012) Stress in Harmonic Serialism. Doctoral dissertation, University of Massachusetts Amherst.
- Ryan, Kevin. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28: 413-454.
- Samek-Lodovici, Vieri and Alan Prince (2005) Fundamental properties of harmonic bounding. Unpublished paper, University of London and Rutgers University, London and New Brunswick, NJ. [Available at: <http://roa.rutgers.edu/>.]
- Smolensky, Paul (2006a) Harmony in linguistic cognition. *Cognitive Science* 30: 779–801.

Smolensky, Paul (2006b) Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. In Paul Smolensky and Géraldine Legendre (eds.), *The harmonic mind: From neural computation to optimality theoretic grammar, vol. 2: Linguistic and philosophical implications* 27–160. Cambridge, MA: MIT Press.

Smolensky, Paul and Legendre, Géraldine (2006). *The harmonic mind: From neural computation to optimality theoretic grammar*. Cambridge, MA: MIT Press.

Staub, Robert (This volume). Serial Restrictions on Feature/Stress Interactions.

Staub, Robert and Joe Pater (This volume). Learning serial constraint-based grammars.

Staub, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John J. McCarthy and Joe Pater (2010) OT-Help 2.0. Software package. Amherst, MA: University of Massachusetts Amherst.

Steriade, Donca (1995) Underspecification and markedness. In John Goldsmith (ed.) *The handbook of phonological theory*. Cambridge, Mass. & Oxford: Blackwell. 114–174.

Steriade, Donca (1999). Phonetics in phonology: the case of laryngeal neutralization. *UCLA Working Papers in Linguistics: Papers in Phonology* 3. 25–145.

Tesar, Bruce and Smolensky, Paul (2000) *Learnability in optimality theory*. Cambridge, MA: MIT Press.

Wilson, Colin (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30: 945–982.

Wilson, Colin (2011) A Targeted Spreading Imperative for Nasal Place Assimilation. In Forthcoming NELS Proceedings.

---

<sup>1</sup> This paper reconfigures and slightly revises portions of Pater (2009a) and Pater (2009b), and adds some new material (especially secs. 3 and 5). Special thanks to Karen Jesney and John McCarthy for extended discussion of most of this work and for helpful comments on the manuscript, and to Paul de Lacy and Chris Potts for discussion of the section 3 research, as well as my other collaborators on Potts et al. (2009) and Staub et al. (2010) – Rajesh Bhatt, Michael Becker, Patrick Pratt and Robert Staub – both projects pushed this work forward considerably. Along with all others thanked in the acknowledgments of the above papers, I particularly appreciate the feedback of participants in our grant group meetings, and in a 2011 LSA summer institute course. This research was supported by NSF Grant 0813829 to the University of Massachusetts Amherst.

<sup>2</sup> In cases of ‘collective bounding’ (Samek-Lodovici and Prince 2005), where a candidate cannot be made optimal because it loses under some weightings or rankings to one candidate, and under some other weightings/rankings to another, whether HG and OT yield the same results depends on the particular violation profiles in question. Anticipating the discussion below, under a symmetric trade-off between constraint violations, OT and HG produce identical results. For example, neither can make [dak.bad] or [dag.bat] optimal for /dag.bad/ in the example in (10). One exception is that Maximum Entropy Grammar can give such collectively bounded candidates non-negligible probability – see section 5 for further discussion. When the trade-off is asymmetric, OT and HG can diverge – see section 4 for examples of candidates that are collectively bounded in OT that are optimal in HG.

<sup>3</sup> I assume a version of OT in which every constraint ranking is a total order, and there is thus only one highest ranked distinguishing constraint. For simplicity, I set up the discussion with the parallel assumption that there is a single highest weighted constraint, but this is not crucial to any of the results.

<sup>4</sup> For an infinite dataset, ranking can produce patterns that elude weighting. For example, the ranking Weight-to-Stress >> Align-R discussed in section 4 produces a pattern in which a heavy syllable will get stressed, no matter how far it is from the right edge of the word, and therefore no

---

matter how many violations of Align-R it incurs. This is impossible to reproduce in HG: regardless of how much higher the weight of Weight-to-Stress is (up to infinity), there will be some distance from the right edge of the word that will make the cost of satisfying that constraint too high relative to Align-R. For directly observable typology, however, it is correct to say that with a given set of constraints, HG will produce a superset of the languages that OT does.

<sup>5</sup> Prince and Smolensky's H-Nuc directly orders candidates rather than simply assigning violation marks in the fashion of \*C-Nuc and maybe all other OT constraints. This allows it to deal with aspects of the Berber syllabification data that \*C-Nuc fails on in a parallel theory. See Pater (2012) for discussion, and for a serial HG analysis using \*C-Nuc that has some advantages over the OT analyses of Berber in Prince and Smolensky (1993/2004).

<sup>6</sup> In terms of the present model, a typological restriction on the size of windows can be obtained by imposing a maximum value on the weights and a minimum difference between the harmony scores of the optima and their competitors (see Boersma and Pater this volume and Potts *et al.* 2010 on the margin of separation of harmony; see relatedly Albright, Magri and Michaels 2008). For example, if we require the optimal candidate's harmony to exceed that of any of its competitors by at least 0.5, then a three-syllable window requires a minimum constraint weight of 2.5, while a four-syllable window requires 3.5. The requirements of a large margin of separation and a maximum value on weights are commonly imposed in learning algorithms in the machine learning and neural modeling literature, and seem quite plausible as biological limitations. For example, in terms of Boersma and Pater's (this volume) Noisy HG, a minimum margin of separation is needed to overcome a given amount of noise. Probabilistic versions of these limits would further increase the learning difficulty for larger stress windows mentioned in the text.

<sup>7</sup> See McCarthy (2009, this volume) for arguments that syllabification must at least sometimes be done in parallel. One way of reconciling those arguments with the serial results of Elfner (this volume) and Pater (2012) would be to have serial initial syllabification, and parallel resyllabification. This could be formalized as a "free" resyllabification operation that applies whenever a segment is stranded by the application of some other operation.

<sup>8</sup> Bane and Riggle (2012) point to other problematic patterns that arise in parallel HG with these sort of basic syllable structure constraints, and at least some of these are resolved by adopting a serial version of the theory (and some are also produced by standard OT with a larger constraint set).

<sup>9</sup> Targeted constraints in HS (Wilson 2011) need not suffer from the indeterminacy problems Pruitt (2012) notes for Lapse constraints and other alternatives to gradient Alignment.