Constraint summation in phonological theory*

Giorgio Magri[1] *and Benjamin Storme*[2]
[1] CNRS, University of Paris 8
[2] University of Lausanne

ABSTRACT

The classical constraints used in phonological theory apply to a single candidate at a time. Yet, some proposals in the phonological literature have enriched the classical constraint toolkit with constraints that instead apply to multiple candidates simultaneously. For instance, Dispersion Theory (Flemming 2002, 2004, 2008) adopts *distinctiveness constraints* that penalize pairs of surface forms which are not sufficiently dispersed. Also, some approaches to paradigm uniformity effects (Kenstowicz 1997; McCarthy 2005) adopt *Optimal Paradigm faithfulness constraints* that penalize pairs of stems in a paradigm which are not sufficiently similar. As a consequence, these approaches need to "lift" the classical constraints from a single candidate to multiple candidates by summing constraint violations across multiple candidates. Is this assumption of constraint summation typologically innocuous? Or do the classical constraints make different typological predictions when they are summed, independently of the presence of distinctiveness or optimal paradigm faithfulness constraints? The answer depends on the underlying model of constraint optimization, namely on

Keywords: *constraint-based phonology, Optimality Theory, Harmonic Grammar, Dispersion Theory, paradigm uniformity, Optimal Paradigms model*

*We would like to thank Alan Prince for very useful feedback. We also would like to thank Edward Flemming for calling our attention to McCarthy (2005) as another application of constraint summation. The research reported in this paper has been supported by a grant from the Agence National de la Recherche (project title: The mathematics of segmental phonotactics) and by a grant from the MIT France Seed Fund (project title: Phonological Typology and Learnability).

how the profiles of constraint violations are ordered to determine the smallest one. Extending an independent result by Prince (2015), this paper characterizes those orderings for which the assumption of constraint summation is typologically innocuous. As a corollary, the typological innocuousness of constraint summation is established within both Optimality Theory and Harmonic Grammar.

1 INTRODUCTION

The classical constraints used in the phonological literature evaluate individual candidate surface realizations of a given underlying form (Prince and Smolensky 1993/2004). Yet, some authors have extended this classical constraint toolkit through constraints that evaluate not a *single* candidate but *multiple* candidates simultaneously. One example is provided by *distinctiveness* constraints in *Dispersion Theory* (DT; Flemming 2002, 2004, 2008), which penalize surface forms which are not sufficiently contrastive. Another example is provided by *Optimal Paradigm* (OP) faithfulness constraints in theories of paradigm uniformity effects such as the *Optimal Paradigms model* (OPM; Kenstowicz 1997; McCarthy 2005), which penalize dissimilarities among surface forms in an inflectional paradigm.

The addition of distinctiveness and OP faithfulness constraints to the classical constraint set raises a subtle technical problem: since classical constraints apply to a single candidate at a time while distinctiveness and OP faithfulness constraints instead apply to multiple candidates simultaneously, the classical constraints need to be “lifted” from individual candidates to tuples of candidates. A natural solution to this problem (and indeed the solution pursued in DT and the OPM) is to lift a classical constraint C from individual candidates to tuples of candidates by summing constraint violations across the candidates in the tuple, as in (1).[1]

[1] In order for this assumption (1) to make sense, the sum on the right-hand side must be finite. Finiteness requires one of two conditions to hold: either the sum has only a finite number of addenda; or else it has an infinite number of addenda but only finitely many of them are different from zero because the con-

(1) **Constraint summation assumption:**

$$C\big(\langle \textit{candidate 1, candidate 2, candidate 3}\ldots\rangle\big) =$$
$$= C(\textit{candidate 1}) + C(\textit{candidate 2}) + C(\textit{candidate 3}) + \ldots$$

To set the stage for the paper, Section 2 reviews the arguments for this constraint summation assumption (1) in DT and the OPM.

The use of distinctiveness constraints to model contrast is a topic of intense debate in the current phonological literature (see for instance Blevins 2004; Boersma and Hamann 2008; Stanton 2017), as is the use of OP faithfulness constraints to capture paradigm uniformity effects (see for instance Albright 2010). This paper contributes to these debates by taking a closer look at a formal consequence of these constraints, namely the assumption (1) that classical constraints be lifted through constraint summation. What are the phonological implications of this assumption? To zoom in on this question, let us suppose that the constraint set contains no distinctiveness or OP faithfulness constraints but only classical constraints. We then have two options. According to the classical approach, we can compute the optimal surface realization of each underlying form individually relative to the original classical constraints. Alternatively, we can compute the optimal surface realizations for all the underlying forms simultaneously

straint C only penalizes finitely many of the candidates considered. As we will see in Section 2, in the case of the OPM, the number of addenda on the right-hand side of (1) is indeed finite because it is controlled by the size of the inflectional paradigm, which is a finite set of forms. For applications of DT to segment inventories, the number of addenda on the right-hand side of (1) is also plausibly finite, because it is controlled by the size of the underlying universal inventory of atomic linguistic sounds, which is plausibly finite. Yet, for applications of DT to strings of segments, the number of addenda on the right-hand side of (1) is controlled by the number of strings, which is infinite unless we can cap their length in some phonologically plausible way. Furthermore, it is unlikely that a constraint would penalize only finitely many candidates in this case, as pointed out to us by Edward Flemming (p.c.). For example, if C is a markedness constraint penalizing voiced stops, it will assign violations to the infinite set of strings containing voiced stops. Similarly, if C is an identity faithfulness constraint for voicing, it will be violated by an infinite number of mappings with voiced stops in the input string corresponding to voiceless stops in the output string. We leave this technical issue open at this stage.

relative to the summed version (1) of the classical constraints. Is it the case that these two approaches lead to the same set of winners, so that constraint summation is innocuous? Equivalently, is it the case that phonological theories that make use of constraint summation, such as DT and the OPM, actually coincide with classical constraint-based phonology when the constraint set consists solely of classical constraints but no distinctiveness or OP faithfulness constraints? Section 3 formalizes this question.

Obviously, the individual constraint violations *C*(*candidate 1*) and *C*(*candidate 2*) cannot be reconstructed from their sum *C*(*candidate 1*) + *C*(*candidate 2*). One might thus intuitively expect that the assumption (1) of constraint summation wipes away much of the information encoded by the classical constraints. If that were indeed the case, phonological frameworks such as DT and the OPM which make use of constraint summation could profoundly alter the typological implications of the classical constraints, possibly leading to pernicious typological predictions. The goal of this paper is to show that this pessimism is unwarranted.

To start, Section 4 focuses on the case of *Optimality Theory* (OT; Prince and Smolensky 1993/2004). In OT, violation profiles are optimized relative to the *lexicographic order* induced by some constraint *ranking*. In the context of OT, the typological innocuousness of the assumption (1) of constraint summation has been established in Prince (2015). Interestingly, we observe that Prince's original proof can be substantially simplified if we reason in terms of violation profiles rather than in terms of *elementary ranking conditions* (ERCs; Prince 2002), as Prince does. The fact that an OT-specific tool like ERCs hinders rather than facilitates the proof suggests that Prince's result must be independent of the specifics of OT and instead follow from some deeper, more general structure. What is this structure?

The statistician Michel Talagrand explains why it is important to pursue this question: "The practitioner [...] is likely to be struggling at any given time with his favorite model of the moment, a model that will typically involve a rather rich and complicated structure. There is a near infinite supply of such models. Fashions come and go, and the importance with which we view any specific model is likely to strongly vary over time. [One should thus] always consider a problem under the minimum structure in which it makes sense. [...]

By following [this advice], one is naturally led to the study of problems with a kind of minimal and intrinsic structure. Besides the fact that it is much easier to find the crux of the matter in a simple structure than in a complicated one, there are not so many really basic structures, so one can hope that they will remain of interest for a very long time." (Talagrand 2014)

Pursuing this insight, Section 5 offers a complete (both necessary and sufficient) characterization of the "minimal structure" needed to guarantee the typological innocuousness of the assumption (1) of constraint summation, namely the structure provided by additive weak orders. This characterization shows that OT's specific choice of the lexicographic order is by no means necessary to ensure the typological innocuousness of constraint summation. As discussed in Section 6, typological innocuousness indeed extends beyond OT to a variety of constraint-based optimization schemes, crucially including optimization schemes based on additive utility functions, as in *Linear OT* (LOT; Keller 2000, 2006) and *Harmonic Grammar* (HG; Legendre *et al.* 1990b,a; Smolensky and Legendre 2006), which have figured prominently in the recent phonological literature (Pater 2009; Potts *et al.* 2010). Section 7 concludes the paper by discussing the implications of this result for the formal foundations of phonological approaches that make use of constraint summation, such as DT and the OPM.

WHY IS CONSTRAINT SUMMATION NEEDED IN PHONOLOGICAL THEORY 2

To set the stage for the paper, this section reviews arguments for the constraint summation assumption (1) in DT and in the OPM.[2] Our presentation stresses the complete formal analogy between the two arguments, despite the fact that they belong to distant corners of phonological theory. The rest of the paper will then take a closer look at the constraint summation assumption (1) motivated in this section.

[2] The reader already familiar with DT and the OPM might want to skip ahead to Section 3.

2.1 *Dispersion Theory*

This section summarizes the argument for constraint summation in DT. The argument has three steps. First, Subsection 2.1.1 reviews Flemming's challenge against classical markedness and faithfulness constraints that look at a single candidate at a time. Second, Subsection 2.1.2 overviews Flemming's proposal that the classical constraint toolkit be enriched with *distinctiveness* constraints that look at multiple candidates simultaneously. Third, Subsection 2.1.3 illustrates how the classical constraints are "lifted" to multiple candidates through constraint summation (1) in order for them to be able to interact with distinctiveness constraints.[3]

2.1.1 Insufficiency of classical markedness and faithfulness constraints

The constraint-based phonological literature assumes two classes of constraints. *Faithfulness* constraints measure the distance or discrepancy between an underlying form and its surface realization. *Markedness* constraints measure the phonotactic ill-formedness of a surface form. Both types of constraints thus look at a single underlying/surface form candidate pair at a time. Flemming (2002, 2004) argues that this classical toolkit is insufficient. We review here one of his arguments, based on the typology of systems of contrasts among voiceless, plain voiced, and prenasalized voiced stops (Flemming 2004, pages 258-263). Many languages contrast voiceless stops [p, t, k] with plain voiced stops [b, d, g] (e.g. French; Tranel 1987). Yet there are also a few languages that prefer having prenasalized voiced stops [^{n}b, ^{n}d, ^{n}g]

[3] The architecture summarized in Subsections 2.1.2–2.1.3 is a simplified version of the architecture proposed in Flemming (2008) (not the earlier one proposed in Flemming 2002). Our presentation is simplified because it confounds Flemming's (2008) three modules into a single one. In fact, we ignore Flemming's orthogonal claim that language specific properties of phonetic realization play a role in the phonology. Hence, we conflate Flemming's *realized inputs* with underlying phonological forms and effectively ignore the "phonetic realization module" which derives the former from the latter. Furthermore, we ignore the distinction between the "inventory selection module" and the "phonotactics module", following Flemming (2017b,a). These simplifications are adopted for simplicity only and they do not affect the scope of our results.

(instead of plain voiced stops) contrast with voiceless stops (e.g. San Juan Colorado Mixtec; Iverson and Salmons 1996). How could such a language be derived with classical markedness and faithfulness constraints?

Obviously, we would need a markedness constraint which penalizes plain voiced stops at the exclusion of prenasalized ones. Let us call this constraint *D. The intuition behind this constraint could be that voicing is harder to sustain in a plain voiced stop than in a prenasalized one because the nasal aperture facilitates voicing by preventing a fast pressure buildup above the glottis (Ohala 1983). We assume that this constraint *D is "counterbalanced" by another markedness constraint $*^{n}$D that instead penalizes prenasalized voiced stops at the exclusion of plain ones. The intuition behind this constraint would be that prenasalized stops are more effortful to produce because they "require rapid raising of the velum to produce oral and nasal phases within the same stop" (Flemming 2004, page 260). Finally, we consider a third markedness constraint *VTV which penalizes voiceless stops in intervocalic position. The constraint set is completed by two faithfulness constraints Ident(voice) and Ident(nas) that protect the underlying specifications for voicing and nasalization, respectively.

For concreteness, let us adopt the OT model of constraint interaction (see Section 4 below for a review of the OT formalism). Tableau (2a) derives the faithful realization of underlying voiceless stops intervocalically. And tableau (2b) derives prenasalization of an underlying plain voiced stop.

(2) a.

/ata/	*D	Ident(voice)	Ident(nas)	$*^{n}$D	*VTV
☞ [ata]					*
[ada]	*!	*			
[a^{n}da]		*!	*	*	

b.

/ada/	*D	Ident(voice)	Ident(nas)	$*^{n}$D	*VTV
[ata]		*!			*
[ada]	*!				
☞ [a^{n}da]			*	*	

In conclusion, we have derived a language like San Juan Colorado Mixtec, where voiceless stops contrast with prenasalized voiced stops but not with plain voiced stops.

However, Flemming notes that no language prefers prenasalized voiced stops to plain voiced stops in a context where voiceless stops are banned. For instance, intervocalic voiceless stops are never repaired through intervocalic prenasalization. The only attested repair is intervocalic voicing (e.g. Tümpisa Shoshone; Dayley 1989). Flemming argues that this fact is difficult to derive with only the faithfulness and markedness constraints made available by classical constraint-based phonology. In fact, as soon as *D is allowed to outrank *ⁿD as in tableaux (2), we derive an unattested pattern of intervocalic prenasalization of voiceless stops. This pattern is derived if *VTV and Ident(voice) are flipped, as in (3).

(3)

/ata/	*VTV	*D	*ⁿD	Ident(voice)	Ident(nas)
[ata]	*!				
[ada]		*!		*	
☞ [aⁿda]			*	*	*

These considerations suggest that our initial attempt at deriving the preference for prenasalized over plain voiced stops in Mixtec through classical markedness constraints is not on the right track. The constraint responsible for this preference cannot be a classical markedness constraint such as the constraint *D proposed above, because that constraint is blind to the presence or absence of a plain voiceless stop. This strategy based on classical markedness thus leads to the incorrect prediction that prenasalized voiced stops are preferred also in the absence of plain voiceless stops, namely that prenasalization can be used as a repair strategy for intervocalic voiceless stops.

2.1.2 Distinctiveness constraints

In order to solve this impasse, Flemming proposes that the preference for prenasalized voiced stops in contexts where voiceless stops are available results from contrast enhancement: the voicing contrast is more distinct in the pair [t]-[ⁿd] than in the pair [t]-[d] (Iverson and Salmons 1996), due to the higher intensity of the periodic part of the speech signal in [ⁿd] than in [d]. In the presence of a voiceless stop, the preference for maximizing contrast can exert its effect and allow for [^{n}d] at the exclusion of plain [d]. But in the absence of voiceless stops, there is no contrast to enhance and thus the markedness of [^{n}d] relative to [d] is the only active force, whereby voiced stops are predicted to be systematically preferred.

Flemming formalizes the preference for more distinct contrasts via *distinctiveness* constraints that penalize pairs of sounds based on their distance along a perceptual scale. In the case at hand, the relevant perceptual scale is the intensity of voicing. Following Flemming's simplifying assumption, suppose that the intensity of voicing is equal to 0 in voiceless stops, to 1 in plain voiced stops, and to 2 in prenasalized stops. Pairs [t]-[d] and [d]-[^{n}d] (but not [t]-[^{n}d]) violate a distinctiveness constraint requiring voicing contrasts corresponding to a distance strictly larger than one unit along the intensity scale. This constraint is denoted MinDist, as in (4).

(4) MinDist:
Assign a violation mark to pairs of surface forms with a voicing contrast corresponding to a distance equal to or smaller than 1 along the scale of voicing intensity.
Penalizes [t]-[d], [d]-[^{n}d]. Does not penalize [t]-[^{n}d].

All three pairs [t]-[^{n}d], [t]-[d], and [d]-[^{n}d] violate a distinctiveness constraint requiring voicing contrasts corresponding to a distance strictly larger than two units along the intensity scale (we ignore this constraint in what follows because it does not distinguish among these three pairs).

Lifting classical constraints through constraint summation 2.1.3

Distinctiveness constraints are formally very different from classical faithfulness and markedness constraints. In fact, classical constraints assign a number of violations to each individual candidate surface realization of a given underlying form. Distinctiveness constraints instead compare tuples of surface realizations of multiple underlying forms. This difference has implications for the architecture of grammar. A classical grammar in the constraint-based literature evaluates the candidates of a single underlying form at a time, as illustrated above with the two separate tableaux (2) for the two underlying forms /ata/ and /ada/. A grammar with distinctiveness constraints instead must evaluate tuples of candidates corresponding to multiple underlying forms. But what about the classical constraints that are now mixed up with the distinctiveness constraints? How can they be "lifted" from individual candidates to tuples of candidates of multiple underlying forms? Flemming makes the natural suggestion that classical faithfulness and markedness constraints be redefined for tuples of candidates by summing their constraint violations across all candidates in the tuple, as anticipated in (1).

Tableau (5) illustrates how distinctiveness constraints and constraint summation of the classical constraints work in DT. We consider again the two underlying forms /ata/ and /ada/. This time, they occur together in the same tableau, rather than heading the two separate tableaux in (2). These two underlying forms have three surface candidates [ata], [ada], and [a^{n}da] each in the classical approach of tableaux (2). In DT, we thus consider $3 \times 3 = 9$ pairs of candidates, listed by row in (5). For instance, row (5d) corresponds to the (impossible) mapping whereby /ata/ is realized as [ada] and /ada/ as [ata].

(5)

/ata/, /ada/	MinDist	Ident(voice)	Ident(nas)	$*^{n}$D	*D	*VTV
a. [ata], [ata]		$*_{d\to t}$				$*_{ata}*_{ata}$
b. [ata], [ada]	$*_{t\text{-}d}$				$*_{d}$	$*_{ata}$
c. [ata], [a^{n}da]			$*_{d\to nd}$	$*_{nd}$	$*_{d}$	$*_{ata}$
d. [ada], [ata]	$*_{t\text{-}d}$	$*_{t\to d}*_{d\to t}$			$*_{d}$	$*_{ata}$
e. [ada], [ada]		$*_{t\to d}$			$*_{d}*_{d}$	
f. [ada], [a^{n}da]	$*_{d\text{-}nd}$	$*_{t\to d}$	$*_{d\to nd}$	$*_{nd}$	$*_{d}*_{nd}$	
g. [a^{n}da], [ata]		$*_{t\to nd}*_{d\to t}$	$*_{t\to nd}$	$*_{nd}$	$*_{d}$	$*_{ata}$
h. [a^{n}da], [ada]	$*_{d\text{-}nd}$	$*_{t\to nd}$	$*_{t\to nd}$	$*_{nd}$	$*_{d}*_{nd}$	
i. [a^{n}da], [a^{n}da]		$*_{t\to nd}$	$*_{t\to nd}*_{d\to nd}$	$*_{nd}*_{nd}$	$*_{nd}*_{nd}$	

The distinctiveness constraint MinDist penalizes the pair of surface forms in (5b), because their consonants sit on the voicing scale at a distance of 1 ([ata]-[ada]), respectively. It does not penalize the pair of surface forms in (5c), because their consonants sit on the voicing scale sufficiently far apart, namely at a distance of 2 ([ata]-[a^{n}da]). And so on. This constraint thus exerts a preference for the prenasalized over the plain voiced stop, although crucially only in the presence of the voiceless stop.

Classical markedness and faithfulness constraints are summed across multiple candidates. For instance, a classical faithfulness constraint such as Ident(voice) assigns two violations to the pair of surface forms in (5d), because it assigns one violation to the mapping of /ata/ to [ada] and another violation to the mapping of /ada/ to [ata] and the two violations are summed together, as prescribed by the constraint summation assumption (1). As another example, a classical markedness constraint such as *VTV assigns two violations to the pair of surface forms in (5a), because it features two instances of the surface form [ata]. And so on. To make it easier to track constraint violations, the specific pairs of output segments (in the case of distinctiveness constraints), single output segments (in the case of classical markedness constraints), and input-output segments (in the case of classical faithfulness constraints) that violate the corresponding constraint are indicated in subscript next to each violation mark.[4]

This approach solves the problem discussed in Subsection 2.1.1: it derives a system contrasting voiceless and prenasalized voiced stops while blocking allophonic prenasalization of voiceless stops. In fact, a system with contrasting voiceless and prenasalized voiced stops is derived if MinDist and Ident(voice) are top ranked: this ranking condition eliminates all options but for the desired option $\langle$[ata], [a^{n}da]$\rangle$ in row (5c). Furthermore, nasalization as a repair to intervocalic voiceless stops is impossible because the three logically possible options that prenasalize intervocalic voiceless /t/ are all harmonically bounded. In fact, the option $\langle$[a^{n}da], [ata]$\rangle$ in row (5g) is

[4] As anticipated in the informal discussion at the beginning of Subsection 2.1.2, Flemming assumes that the MinDist distinctiveness constraint is the only constraint that favors prenasalized over plain voiced stops, while classical markedness constraints prefer plain over prenasalized voiced stops. The markedness constraint *D thus needs to be redefined as penalizing all voiced stops, both plain and prenasalized ones. It therefore assigns two violations in (5f), because its two surface forms [ada] and [a^{n}da] both violate it.

harmonically bounded by ⟨[ata], [a^{n}da]⟩ in row (5c). And the options ⟨[a^{n}da], [ada]⟩ and ⟨[a^{n}da], [a^{n}da]⟩ in rows (5h) and (5i) are both harmonically bounded by ⟨[ada], [ada]⟩ in row (5e).

2.2 *Optimal Paradigms model*

This section summarizes the argument for constraint summation in the OPM. The argument has three steps, in complete analogy with the preceding Subsection 2.1. First, Subsection 2.2.1 reviews McCarthy's 2005 challenge that some inflectional paradigms raise for asymmetric, base-prioritizing theories of output-output correspondence. Second, Subsection 2.2.2 overviews McCarthy's proposal that the classical constraint toolkit be enriched with OP faithfulness constraints that evaluate all paradigm members simultaneously. Third, Subsection 2.2.3 illustrates how the classical constraints are "lifted" to entire paradigms through constraint summation (1) in order for them to be able to interact with OP faithfulness constraints.

2.2.1 Insufficiency of asymmetric output-output faithfulness constraints

Morphologically-related forms may bear resemblance that goes beyond what is predicted by the interaction of classical markedness constraints and input-output faithfulness constraints. A classical example is the case of the participle *lightening* [laɪtn̩ɪŋ], where the stem-final consonant is realized as a syllabic nasal [n̩], as in the verb *lighten* [laɪtn̩], instead of the phonotactically expected [n]. Data of this kind have motivated positing another type of faithfulness besides input-output faithfulness: output-output faithfulness.[5] Output-output faithfulness constraints enforce similarity among surface forms. In the case of surface inflected forms, similarity is enforced among surface forms in the same inflectional paradigm, i.e. forms that share a lexeme. In this approach, the presence of syllabic [n̩] in the participle *lightening* [laɪtn̩ɪŋ] can be explained as the result of an output-output faithfulness constraint requiring similarity with the verb *lighten* [laɪtn̩] and

[5] Output-output faithfulness is also motivated by patterns of reduplication (McCarthy and Prince 1995).

outranking the input-output faithfulness constraint requiring similarity with the input stem /laɪtn/.

When one form in a paradigm is morphologically simpler than other paradigm members, this morphologically simpler form (i.e. the base) is always the one that other paradigm members must be faithful to (see Benua 1997, 240–242 for a discussion of potential counterexamples). *Lightening* conforms to this generalization because it is asymmetrically influenced by its base *lighten*. In line with this generalization, theories of output-output faithfulness have been developed where the phonology of the base is computed in a first step and serves as input to the evaluation of affixed forms, alongside the affixed forms' underlying representations (e.g. Benua's 1997 *Transderivational Correspondence Theory*).

However, McCarthy (2005) notes that effects that can be analyzed as output-output faithfulness are also observed in paradigms where all forms are equally complex morphologically and where the choice of the attractor is not guided by morphological simplicity or markedness but by phonological markedness. McCarthy (2005) illustrates his argument with Arabic verbal stems. In Arabic, verbal stems are required to end in VC (e.g. [faʕal], [faʕʕal]). No stem ending in Vːc or VCC is attested in verbal paradigms (e.g. *[faʕaːl], *[faʕl]). This contrasts with nominal stems, which can end in VC, VːC, and VCC (e.g. [faʕal], [faʕaːl], [faʕl]). Under Richness of the Base, the fact that the phonological shape of verbal stems is more constrained than that of nominal stems is unexpected.

McCarthy's insight is that this apparent quirk of Arabic verbs can be explained as an effect of output-output faithfulness, combined with an independent property that distinguishes nouns and verbs in Arabic. Nominal suffixes in Arabic all start with a vowel whereas verbal suffixes start with vowels or consonants (see McCarthy 2005, 179-180 for a list of suffixes). In a nutshell, due to a high ranking markedness constraint banning super heavy syllables (*VːCCV, *VCCCV), verbal stems followed by consonant-initial suffixes can only afford short vowels in stem-final syllables (e.g. [faʕal-tu] but *[faʕaːl-tu] and *[faʕl-tu]). Output-output faithfulness then extends the short vowel that is phonotactically expected before consonant-initial suffixes to the whole paradigm, including to forms built with vowel-initial suffixes and where short vowels are not phonotactically required. In nouns, only vowel-initial suffixes are attested. Therefore, contrary to inflected verbs, there is no paradigmatic pressure to extend stems ending in VC-, therefore allowing for all VC-, V:C-, and VCC- to surface faithfully in inflected nouns.

For this analysis to be implemented using Benua's Transderivational Correspondence Theory, it is necessary to assume that the base in verbal paradigms is one of the forms built with a consonant-initial suffix. Faithfulness to the base then extends the short vowel that is phonotactically expected in this form to all other forms. Tableaux (6a) and (6b) show how this analysis works, focusing on two forms of the paradigm of hypothetical underlying /faʕa:l/: (i) an inflected form with a consonant-initial suffix that serves as the base in the paradigm, /faʕa:l-tu/ (1st singular perfective), and (ii) an inflected form with a vowel-initial suffix, /faʕa:l-a/ (3d singular perfective).

(6) a.

/faʕa:l-tu/	*V:CCV	IdBD(length)	IdIO(length)
[faʕa:ltu]	*		
☞ [faʕaltu]			*

b.

/faʕa:l-a/ Base = [faʕaltu]	*V:CCV	IdBD(length)	IdIO(length)
[faʕa:la]		*	
☞ [faʕala]			*

To get the short vowel to be extended to other forms, /faʕa:l-tu/ has to be considered as the base. As the base, its phonology is computed first. In this first cycle, only input-output correspondence is relevant. Because *V:CCV outranks the faithfulness constraint protecting underlying vowel length (IdentIO(length)), the stem long vowel is shortened before CC, as shown in tableau (6a). In a second step, the phonology of other paradigm members is computed. Now, output-output correspondence is relevant, with [faʕa:ltu] serving as the base for the surface form derived from underlying /faʕa:l-a/. IdentBD(length) requires the form under evaluation to match the base along vowel length. This constraint outranks IdentIO(length), therefore favoring base-derivative similarity over input-output similarity, as shown in tableau (6b).

As pointed out by McCarthy, the problem with this approach is that there is no independent, morphological motivation to treat a form with a consonant-initial suffix (/faʕaːl-tu/ in our case) as the base. Indeed, inflected forms with consonant-initial suffixes are neither simpler than the others paradigm members (all forms are inflected) nor morphosyntactically less marked. Indeed, morphosyntactic markedness predicts that the third person singular form should be the base. However, all third person singular forms in the verbal paradigm are built with vowel-initial suffixes (cf. /faʕaːl-a/ in our example).

Optimal Paradigms faithfulness constraints 2.2.2

To solve this issue, McCarthy proposes *Optimal Paradigm* faithfulness constraints. Surface inflected forms are related by output-output correspondence to all other inflected forms of the same stem. The stem of every paradigm member stands in correspondence with the stem of other members; OP faithfulness constraints enforce similarity among corresponding stems in a paradigm. The resulting system is distinct from Benua's Transderivational Correspondence Theory because the latter is asymmetrical (the base is generated "first", hence not modifiable) while the effects of OP faithfulness are symmetric: all members in a paradigm are evaluated simultaneously hence each of them can be modified.

In the case of Arabic, extension of the short vowel from stems built with consonant-initial suffixes is enforced by an OP faithfulness constraint that requires matching vowel length in all pairs of paradigm members, as defined in (7).

(7) Ident-OP(length)
In every paradigm, the stem of each paradigm member corresponds to the stem of every other paradigm member.
In each pair of correspondent stems S1-S2, assign a violation mark for each vowel in S1 that does not have the same length as the corresponding vowel in S2.
Penalizes paradigm <faʕaːl-ta, faʕal-u> and <faʕal-ta, faʕaːl-u>. Does not penalize paradigms <faʕal-ta, faʕal-u> and <faʕaːl-ta, faʕaːl-u>

This constraint evaluates surface resemblance *symmetrically* across inflectionally related forms, hence it does not stipulate that any paradigm member should be a priori preferred over the others. In a concrete analysis, the choice of the attractor is determined by markedness, as will be shown in more detail in the next subsection.

2.2.3 Lifting classical constraints through constraint summation

OP faithfulness constraints are formally very different from classical faithfulness and markedness constraints. In fact, classical constraints assign a number of violations to each individual candidate surface realization of a given underlying form. OP faithfulness constraints instead compare the surface realizations of multiple underlying forms based on their similarity. Again as in the case of DT discussed in the preceding subsection, this difference means that classical constraints need to be "lifted" from individual candidates to whole paradigms in order to be able to interact with OP faithfulness constraints. McCarthy (2005, p.173) makes the natural suggestion that classical faithfulness and markedness constraints be redefined by summing their constraint violations across all forms in a paradigm, as anticipated in (1).

Tableau (8) illustrates how OP faithfulness constraints and constraint summation of the classical constraints work in the OPM. We consider again the two underlying forms /faʕaːl-a/ and /faʕaːl-tu/. This time, they occur together in the same tableau, rather than heading the two separate tableaux in (6). These two underlying forms have two surface candidates each in the classical approach of tableaux (6). In the OPM, we thus consider $2 \times 2 = 4$ pairs of candidates, listed by row in (8). For instance, row (8a) corresponds to the mapping whereby /faʕaːl-a/ is realized as [faʕaːl-a] and /faʕaːl-tu/ as [faʕaːl-tu].

(8)

/faʕaːl-a/, /faʕaːl-tu/	*V:CCV	IdentOP(length)	IdentIO(length)
a. [faʕaːl-a], [faʕaːl-tu]	*		
b. [faʕaːl-a], [faʕal-tu]		*	*
c. [faʕal-a], [faʕaːl-tu]	*	*	*
☞ d. [faʕal-a], [faʕal-tu]			**

Classical markedness and faithfulness constraints are summed across multiple candidates. For instance, a classical faithfulness constraint such as IdentIO(length) assigns two violations to the pair of surface forms in row (8d), because it assigns one violation to the mapping of /faʕaːl-a/ to [faʕaːl-a] and another violation to the mapping of /faʕaːl-tu/ to [faʕaːl-tu] and the two violations are summed together, as prescribed by the constraint summation assumption (1).

The OP faithfulness constraint IdentOP(length) penalizes the pairs of surface forms in (8b) and (8c), because they feature two vowels that stand in correspondence but do not match in length. This constraint thus exerts a preference for paradigm uniformity, without specifying which form will be the attractor: the pairs of surface forms in (8a) and (8d) are equally good in terms of paradigm uniformity. The choice of the attractor is determined by the high ranked markedness constraint *VːCCV, which penalizes (8a) featuring a super heavy syllable. This approach solves the problem discussed in Subsection 2.2.1: it derives the generalization on verbal stems without needing to stipulate that inflected forms with consonant-initial suffixes are the base, in the morphological sense of Benua (1997). The reason why the short vowel length is extended to other forms rather than the long one is the fact that super heavy syllables are marked.

IS CONSTRAINT SUMMATION TYPOLOGICALLY INNOCUOUS? 3

The preceding section has reviewed some phonological theories (such as DT and the OPM) that share two formal innovations. The first innovation is that the classical constraint set is enriched with constraints (such as distinctiveness and OP faithfulness constraints) that evaluate multiple candidates *simultaneously* by comparing surface forms from the perspective of their distinctiveness or their mutual faithfulness. These constraints are therefore formally rather different from classical faithfulness and markedness constraints, which instead evaluate candidates *individually*, one at a time. The second related innovation is that classical markedness and faithfulness constraints are "lifted" from an individual candidate to multiple candidates through the constraint summation assumption (1). Is this constraint summation assumption *typologically innocuous?* In other words, is it the case that constraint summation does not alter the typological implications of classical markedness and faithfulness constraints? Or is it instead the case that phonological theories (such as DT and the OPM) that make use of constraint summation predict very different typologies even when the constraint set consists only of classical constraints (but no distinctiveness or OP faithfulness constraints)? This section formulates this question explicitly.

3.1 *The classical approach*

Let us suppose that we have only two underlying forms. The reasoning developed in this and the following sections extends straightforwardly from two to an arbitrary finite number of underlying forms (see footnote 7 below; the extension to an infinite number of underlying forms is trickier, as discussed in footnote 1 above). In order to focus on the constraint summation assumption (1), we suppose that the constraint set consists of n classical constraints $C_1, \ldots, C_n$, but no distinctiveness or OP faithfulness constraints. We denote by α the generic surface candidate of the first underlying form and we collect these surface candidates into a candidate set A. The classical constraints $C_1, \ldots, C_n$ assign to (the mapping of that underlying form into) the candidate α the n constraint violations $a_1, \ldots, a_n$. We collect them together into a tuple $\mathbf{a} = (a_1, \ldots, a_n)$. Analogously, we denote by β the generic surface candidate of the second underlying form and we collect these surface candidates into a candidate set B. The classical constraints $C_1, \ldots, C_n$ assign to (the mapping of that second underlying form into) the candidate β a tuple $\mathbf{b} = (b_1, \ldots, b_n)$ of n constraint violations $b_1, \ldots, b_n$. A concrete example of the sets A and B is provided by the two tableaux (2a) and (2b) for the two underlying forms /ata/ and /ada/. In this case, $n = 5$, the candidate corresponding to the first row of the tableau A is $\alpha =$ [ata], and the corresponding tuple of constraint violations is $\mathbf{a} = (0, 0, 0, 0, 1)$.

Under the assumption that the constraints suffice to capture all the relevant information, the optimal candidate for a given underlying form must be the one which violates the constraints the least, that is which corresponds to the "smallest" tuple of constraint violations. To formalize this intuition, we extend the intuitive notion of "smaller than" from single numbers to tuples of numbers. Thus, the condition $\widehat{\mathbf{a}} < \mathbf{a}$ means that the tuple of constraint violations $\widehat{\mathbf{a}}$ is smaller than the tuple of constraint violations $\mathbf{a}$. Effectively, this means that we define an *order* $<$ among tuples of constraint violations. Different implementations of (classical) constraint-based phonology considered in the literature differ for the choice of the order $<$ used to compare tuples of constraint violations. For full generality, we allow this order $<$ to be *partial*: for some pairs of tuples of constraint violations, $<$ might not be able to tell which one is smaller. In other words, some tuples might be *incommensurable*.[6]

We denote by $\mathrm{opt}_< A$ the collection of *optimal* candidates in the set A, namely those candidates $\widehat{\alpha}$ corresponding to a tuple $\widehat{\mathbf{a}}$ of constraint violations which is minimal relative to the order $<$, as defined in (9).

(9) $\mathrm{opt}_< A$ is the set of those candidates $\widehat{\alpha}$ in A such that there exists no competing candidate α in A such that the constraints assign to this competing candidate α a tuple $\mathbf{a}$ of constraint violations which is smaller than the tuple $\widehat{\mathbf{a}}$ of constraint violations they assign to the optimal candidate $\widehat{\alpha}$, namely $\mathbf{a} < \widehat{\mathbf{a}}$.

(Classical) constraint-based phonology assumes that the underlying form with candidate set A is mapped to a surface candidate $\widehat{\alpha}$ which violates the constraints the least, namely which belongs to the optimal subset $\mathrm{opt}_< A$ of candidates with the smallest tuples of constraint violations. Analogous considerations hold for the other underlying form, which is mapped to a surface candidate in the optimal set $\mathrm{opt}_< B$.

We note that the set $\mathrm{opt}_< A$ of optimal candidates can contain more than one candidate. In fact, two candidates $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ can both be optimal if their corresponding tuples of constraint violations $\widehat{\mathbf{a}}_1$ and $\widehat{\mathbf{a}}_2$ are incommensurable: neither of the two is larger than the other according to the order $<$ because $<$ is only partial. Furthermore, even

[6] For instance, in the case of the HG implementation of constraint-based phonology (see Subsection 6.2 below), the order $<$ is defined in terms of a *utility* or *harmony function*. Two candidates with different tuples of constraint violations can achieve the same harmony. Their tuples of constraint violations are therefore incommensurable relative to this order $<$.

if $<$ is total, two different candidates $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ can both be optimal if the constraints fail to distinguish between them, namely the two candidates share the same tuple $\widehat{\mathbf{a}}_1 = \widehat{\mathbf{a}}_2$ of constraint violations, as we will discuss in more detail in Subsection 3.4.

3.2 *The constraint summation approach of DT and the OPM*

We denote by $A \times B$ the collection of all pairs (α, β) of a candidate α from the candidate set A and a candidate β from the candidate set B. We "lift" the n classical constraints from single candidates to pairs of candidates through the constraint summation assumption (1). This means that the lifted constraints assign to the candidate pair (α, β) the component-wise sum $\mathbf{a}+\mathbf{b}$ of the tuples $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ of constraint violations assigned to the two individual candidates α and β, as in (10).

(10) $\mathbf{a}+\mathbf{b} = (a_1 + b_1, \ldots, a_k + b_k, \ldots, a_n + b_n)$

To illustrate, if A and B are the two tableaux (2a) and (2b), their product $A \times B$ is the tableau (11), which lists all pairs of candidates and sums the stars in the two corresponding cells of A and B.

(11)

	(/ata/, /ada/)	Ident(voice)	Ident(nas)	$^*{}^{\mathrm{n}}$D	*D	*VTV
a.	([ata], [ata])	*				**
b.	([ata], [ada])				*	*
c.	([ata], [$\mathrm{a^n}$da])		*	*	*	*
d.	([ada], [ata])	**			*	*
e.	([ada], [ada])	*			**	
f.	([ada], [$\mathrm{a^n}$da])	*	*	*	**	
g.	([$\mathrm{a^n}$da], [ata])	**	*	*	*	*
h.	([$\mathrm{a^n}$da], [ada])	*	*	*	**	
i.	([$\mathrm{a^n}$da], [$\mathrm{a^n}$da])	*	**	**	**	

This is of course the same as tableau (5) considered above, stripped of the column corresponding to the distinctiveness constraint MinDist.

We denote by $\mathrm{opt}_<(A \times B)$ the collection of *optimal* candidate pairs in the set $A \times B$, namely those candidate pairs $(\widehat{\alpha}, \widehat{\beta})$ such that the tuples $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{b}}$ of constraint violations corresponding to the candidates $\widehat{\alpha}$ and $\widehat{\beta}$ yield a minimal sum $\widehat{\mathbf{a}} + \widehat{\mathbf{b}}$, as defined in (12). This is of course a special case of the definition (9) of optimality, applied to the set $A \times B$ with summed tuples of constraint violations rather than to the set A with the original tuples of constraint violations.

(12) $\mathrm{opt}_<(A \times B)$ is the set of those candidate pairs $(\widehat{\alpha}, \widehat{\beta})$ in $A \times B$ such that there exists no competing candidate pair (α, β) in $A \times B$ such that the sum $\mathbf{a} + \mathbf{b}$ of the two tuples $\mathbf{a}$ and $\mathbf{b}$ of constraint violations assigned to the two candidates α and β is smaller than the sum $\widehat{\mathbf{a}} + \widehat{\mathbf{b}}$ of the two tuples $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{b}}$ of constraint violations assigned to the two candidates $\widehat{\alpha}$ and $\widehat{\beta}$, namely $\mathbf{a} + \mathbf{b} < \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$.

Phonological theories which make use of constraint summation, such as DT and the OPM, assume that the two underlying forms considered here are mapped to the pair of surface candidates $(\widehat{\alpha}, \widehat{\beta})$ which violates the constraints the least, namely which belongs to the optimal set $\mathrm{opt}_<(A \times B)$ of candidate pairs with the smallest summed tuple of constraint violations.

Typological innocuousness as a commutativity identity 3.3

Let us take stock. According to the classical implementation of constraint-based phonology reviewed in Subsection 3.1, the order $<$ is used twice. It is used once to assign to the first underlying form a candidate $\widehat{\alpha}$ in the set $\mathrm{opt}_< A$ of optimal candidates of A. It is then used again and independently to assign to the second underlying form a candidate $\widehat{\beta}$ in the set $\mathrm{opt}_< B$ of optimal candidates of B. According to the constraint summation approach of DT and the OPM reviewed in Subsection 3.2, the order $<$ is instead used only once to assign to the two underlying forms considered simultaneously a candidate pair $(\widehat{\alpha}, \widehat{\beta})$ in the set $\mathrm{opt}_<(A \times B)$ of optimal pairs of $A \times B$, where candidate pairs are compared based on the sums of constraint violations of the two individual candidates, by virtue of the constraint summation assumption (1).

Suppose now that a candidate pair is optimal iff it consists of two optimal candidates, as stated in (13). This means that the two underlying forms considered end up with the same optimal candidates no matter whether we adopt the classical approach or the approach based on constraint summation of DT and the OPM. In other words, the constraint summation assumption (1) made by DT and the OPM would be *typologically innocuous*. And classical constraint-based phonology would thus follow as a special case of DT and the OPM when the constraint set contains no distinctiveness or OP faithfulness constraints.[7]

(13) **Commutativity identity:**

$$\underbrace{\operatorname{opt}_{<}(A \times B)}_{\text{constraint summation approach (DT/OPM)}} = \underbrace{\operatorname{opt}_{<} A \times \operatorname{opt}_{<} B}_{\text{classical approach}}$$

In conclusion, a crucial issue of the formal analysis of phonological theories such as DT and the OPM is whether the identity (13) holds in the general case, for any two candidate sets A and B. In other words, whether the two operations of optimization and product *commute*: by first combining (through $\times$) candidates from A and B into pairs and then optimizing (through $\operatorname{opt}_<$) over candidate pairs relative to the summed constraint violations (as prescribed by the left-hand side, which corresponds to the summation based approach of DT or the OPM) we get the same result that we get by first optimizing (through $\operatorname{opt}_<$) within the two separate candidate sets A and B and then combining (through $\times$) optimal candidates into pairs (as prescribed by the right-hand side, which corresponds to the classical approach in constraint-based phonology).

[7] As anticipated at the beginning of this section, the discussion extends straightforwardly from the case of only two underlying forms considered here to the case of an arbitrary finite number of underlying forms. Indeed, suppose we have three underlying forms with candidate sets A, B, and C. The commutativity identity that we need to establish in this case is $\operatorname{opt}_<(A \times B \times C) = \operatorname{opt}_< A \times \operatorname{opt}_< B \times \operatorname{opt}_< C$. The latter follows by applying (13) twice: once to the two sets $A \cup B$ and C, to ensure that $\operatorname{opt}_<(A \times B \times C) = \operatorname{opt}_<(A \times B) \times \operatorname{opt}_< C$; then again to the two sets A and B, to ensure that $\operatorname{opt}_<(A \times B) = \operatorname{opt}_< A \times \operatorname{opt}_< B$.

Constraint distinctiveness is not preserved by constraint summation 3.4

The sum $\mathbf{a} + \mathbf{b}$ of two tuples of constraint violations carries less information than the two individual tuples $\mathbf{a}$ and $\mathbf{b}$: the summed tuple is computed from the two individual tuples but the individual tuples cannot be univocally reconstructed from the summed tuple. The assumption (1) of constraint summation can thus wipe away potentially crucial information encoded in the individual tuples of constraint violations, imperiling the validity of the commutativity identity (13). To appreciate the problem, let us look at the behavior of constraint distinctiveness under constraint summation.

Suppose that the (classical) constraints $C_1, \ldots, C_n$ considered are *distinctive*. This means that any two candidates in the candidate set A and any two candidates in the candidate set B are distinguished by at least one constraint. Equivalently, no two candidates in A and no two candidates in B are assigned identical tuples of constraint violations. Suppose furthermore that the order $<$ over tuples of constraint violations is *total*: any two different tuples are ordered relative to each other. Distinctiveness and totality together ensure that the set $\text{opt}_< A$ of optimal candidates of A and the set $\text{opt}_< B$ of optimal candidates of B are both singleton sets. Their product $\text{opt}_< A \times \text{opt}_< B$ on the right-hand side of the commutativity identity (13) therefore consists of a single candidate pair. The commutativity identity thus requires that also the set $\text{opt}_< (A \times B)$ of optimal candidate pairs in the product $A \times B$ consists of a single pair. But the assumption that $<$ is a total order does not suffice to ensure that, because $A \times B$ could contain two different pairs of candidates which share the same summed tuple of constraint violations, despite the individual candidates in A and B all having distinct tuples of constraint violations. In other words, distinctiveness can be lost when constraint violations are added together.

As a concrete example, consider the two candidate sets A and B described by the two tableaux (2). The tuples of constraint violations listed there are all distinct. If the order $<$ is total, the two sets $\text{opt}_< A$ and $\text{opt}_< B$ of optimal candidates are thus each a singleton. And their product $\text{opt}_< A \times \text{opt}_< B$ on the right-hand side of the commutativity identity (13) thus consists of a single candidate pair. Yet, the product $A \times B$ contains the two different candidate pairs ([ada], [a$^{\text{n}}$da]) and ([a^{n}da], [ada]) whose tuples of summed constraint violations are identical, as shown in (11f) and (11h). Suppose that the order $<$ is defined in such a way that this shared summed tuple happens to be minimal relative to the total order $<$. This means that the set $\mathrm{opt}_{<}(A \times B)$ of optimal candidate pairs in $A \times B$ contains both pairs ([ada], [a^{n}da]) and ([a^{n}da], [ada]). The commutativity identity (13) thus fails, because its right-hand side is a singleton while its left-hand side is not.

These considerations show that we have every reason to expect the commutativity identity (13) to fail in the general case, whereby classical constraint-based phonology cannot be construed as a special case of DT and the OPM, even when the constraint set contains no distinctiveness or OP faithfulness constraints. Can we nonetheless isolate and characterize some special class of orders $<$ among tuples of constraint violations whose special properties validate the commutativity identity (13)? This is the question that we will tackle and solve in the rest of the paper.

4 CONSTRAINT SUMMATION IS TYPOLOGICALLY INNOCUOUS IN OT: PRINCE (2015)

In Section 2, we have reviewed some approaches to phonology (such as DT and the OPM) that assume that classical faithfulness and markedness constraints are summed across multiple candidates, as stated in (1). In Section 3, we have formalized the question of the typological innocuousness of constraint summation through the commutativity identity (13). In this section, we review a result by Prince (2015) showing that this commutativity identity indeed holds in OT. In other words, despite constraint summation, theories such as DT and the OPM make the same typological predictions as classical OT when the constraint set only consists of classical constraints and no distinctiveness or OP faithfulness constraints, whereby constraint summation is typologically innocuous. The next section will then extend this result beyond OT.

Prince's (2015) result 4.1

Prince (2015) focuses on the special case where the order $<$ over tuples of constraint violations is OT's lexicographic order. Let us recall here the explicit definition of this order, that we have already used implicitly in Section 2. We start by linearly ordering or *ranking* the n constraints $C_1, C_2, \ldots, C_n$ in some arbitrary way. Without loss of generality, we assume that constraint C_1 is ranked at the top, constraint C_2 is ranked right underneath it, and so on. The inequality $\mathbf{a} < \widehat{\mathbf{a}}$ then holds between any two tuples of constraint violations $\mathbf{a} = (a_1, \ldots, a_n)$ and $\widehat{\mathbf{a}} = (\widehat{a}_1, \ldots, \widehat{a}_n)$ provided there exists some integer k between 1 and n which validates the conditions in (14).

(14)
$$\begin{array}{rcl} a_1 & = & \widehat{a}_1 \\ & \vdots & \\ a_{k-1} & = & \widehat{a}_{k-1} \\ a_k & < & \widehat{a}_k \end{array}$$

These conditions say that the $k-1$ top ranked constraints assign the same number of violations to the two candidates corresponding to the tuples $\mathbf{a}$ and $\widehat{\mathbf{a}}$.[8] And that the kth constraint is then *decisive* because it assigns less violations to the candidate corresponding to the tuple $\mathbf{a}$ than to the candidate corresponding to the tuple $\widehat{\mathbf{a}}$. Constraints ranked underneath play no role. In Section 5, we will consider alternative ways of ordering tuples of constraint violations.

Prince's result, rephrased below as Proposition 1, says that no ranking information is lost by summing together constraint violations in the case of OT, in the sense that the commutativity identity (13) holds for any candidate sets. In other words, when the constraint set consists of classical constraints only, theories which use constraint summation (such as DT and the OPM) coincide with classical OT and constraint summation is therefore typologically innocuous.

PROPOSITION 1 **(Prince 2015)** *The commutativity identity (13) holds for any two candidate sets A and B relative to OT's lexicographic order $<$ corresponding to any constraint ranking: a candidate $\widehat{\alpha}$ belongs to the set* $\mathrm{opt}_{<} A$ *of OT optimal candidates of A and a candidate $\widehat{\beta}$ belongs to*

[8] These conditions are interpreted as vacuously true if $k = 1$.

the set $\text{opt}_{<} B$ *of OT optimal candidates of B if and only if the candidate pair* $(\widehat{\alpha}, \widehat{\beta})$ *belongs to the set* $\text{opt}_{<}(A \times B)$ *of optimal candidate pairs in* $A \times B$*, when candidate pairs are compared based on summed constraint violations.*

4.2 *A simple proof of Prince's result*

Prince proves Proposition 1 using a piece of notation specifically tailored to OT, namely *elementary ranking conditions* (ERCs; Prince 2002). But this line of reasoning turns out to be involved, intuitively because the operation of constraint summation does not admit a simple counterpart in the theory of ERCs. Yet, Proposition 1 admits an elementary explanation when we reason directly in terms of violation profiles rather than ERCs. In order to streamline the proof of the proposition, we split the commutativity identity (13) into the two inclusions (15) and consider them separately.

(15) a. $\underset{<}{\text{opt}}(A \times B) \subseteq \underset{<}{\text{opt}}\, A \times \underset{<}{\text{opt}}\, B$

b. $\underset{<}{\text{opt}}(A \times B) \supseteq \underset{<}{\text{opt}}\, A \times \underset{<}{\text{opt}}\, B$

To establish the inclusion (15a), let us assume by contradiction that it fails. This means that the candidate pair $(\widehat{\alpha}, \widehat{\beta})$ is OT optimal in $A \times B$ but that, say, the candidate $\widehat{\alpha}$ is not OT optimal in A. This contradictory assumption means that there exists a different candidate α in A that beats (has smaller constraint violations than) candidate $\widehat{\alpha}$. In other words, the tuples $\mathbf{a} = (a_1, \ldots, a_n)$ and $\widehat{\mathbf{a}} = (\widehat{a}_1, \ldots, \widehat{a}_n)$ of constraint violations of the two candidates α and $\widehat{\alpha}$ satisfy the inequality $\mathbf{a} < \widehat{\mathbf{a}}$. This inequality says that there exists $k \in \{1, \ldots, n\}$ such that conditions (14) hold. By adding the corresponding components $\widehat{b}_1, \ldots, \widehat{b}_{k-1}, \widehat{b}_k$ of the tuple $\widehat{\mathbf{b}}$ of constraint violations of candidate $\widehat{\beta}$ to both sides of the inequalities (14), we obtain (16).

(16)
$$\begin{aligned} a_1 + \widehat{b}_1 &= \widehat{a}_1 + \widehat{b}_1 \\ &\vdots \\ a_{k-1} + \widehat{b}_{k-1} &= \widehat{a}_{k-1} + \widehat{b}_{k-1} \\ a_k + \widehat{b}_k &< \widehat{a}_k + \widehat{b}_k \end{aligned}$$

Conditions (16) say that $\mathbf{a} + \widehat{\mathbf{b}} < \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$. In other words, the candidate pair $(\alpha, \widehat{\beta})$ beats the candidate pair $(\widehat{\alpha}, \widehat{\beta})$. This conclusion contradicts the assumption that the candidate pair $(\widehat{\alpha}, \widehat{\beta})$ is OT optimal in $A \times B$.

The proof of the reverse inclusion (15b) is analogous. Indeed, let us assume by contradiction that the candidate $\widehat{\alpha}$ is OT optimal in A and that the candidate $\widehat{\beta}$ is OT optimal in B but that the pair $(\widehat{\alpha}, \widehat{\beta})$ is not OT optimal in $A \times B$. This means that there exists a different pair (α, β) in $A \times B$ such that $\mathbf{a} + \mathbf{b} < \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$, where $\mathbf{a}, \mathbf{b}, \widehat{\mathbf{a}}, \widehat{\mathbf{b}}$ are the tuples of constraint violations of the four candidates $\alpha, \beta, \widehat{\alpha}, \widehat{\beta}$. Suppose that $\mathbf{a} \neq \widehat{\mathbf{a}}$. Since the lexicographic order $<$ is total and $\widehat{\alpha}$ is optimal in A, then $\widehat{\mathbf{a}} < \mathbf{a}$. This means that there exists h such that $\widehat{a}_1 = a_1, \ldots, \widehat{a}_{h-1} = a_{h-1}, \widehat{a}_h < a_h$. Analogously, suppose that $\mathbf{b} \neq \widehat{\mathbf{b}}$. Again, since $<$ is a total order and $\widehat{\beta}$ is optimal in B, then $\widehat{\mathbf{b}} < \mathbf{b}$. This means that there exists k such that $\widehat{b}_1 = b_1, \ldots, \widehat{b}_{k-1} = b_{k-1}, \widehat{b}_k < b_k$. Suppose without loss of generality that $h \geq k$. Thus $\widehat{a}_1 + \widehat{b}_1 = a_1 + b_1, \ldots, \widehat{a}_{k-1} + \widehat{b}_{k-1} = a_{k-1} + b_{k-1}, \widehat{a}_k + \widehat{b}_k < a_k + b_k$. This means that $\widehat{\mathbf{a}} + \widehat{\mathbf{b}} < \mathbf{a} + \mathbf{b}$, contradicting the assumption $\mathbf{a} + \mathbf{b} < \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$. The cases where either $\mathbf{a} = \widehat{\mathbf{a}}$ or $\mathbf{b} = \widehat{\mathbf{b}}$ are treated analogously.

Back to the issue of constraint distinctiveness 4.3

Having understood the reasoning behind Prince's Proposition 1, let us now go back to the issue of constraint distinctiveness discussed in Subsection 3.4. We suppose that the candidate sets A and B are distinctive: no two candidates in A and no two candidates in B share the same tuple of constraint violations. Since OT's lexicographic order $<$ is total, the corresponding sets $\text{opt}_< A$ and $\text{opt}_< B$ of OT optimal candidates are both singletons. Their product $\text{opt}_< A \times \text{opt}_< B$ thus consists of a single pair. Yet, there can exist two different candidate pairs (α, β) and $(\widehat{\alpha}, \widehat{\beta})$ in $A \times B$ which share the same summed tuple $\mathbf{a} + \mathbf{b} = \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$ of constraint violations, because distinctiveness of the individual candidate sets A and B does not entail distinctiveness of their product $A \times B$ when the constraint violations of a pair of candidates are obtained by summing together the constraint violations of the two individual candidates. The assumption that OT's lexicographic order $<$ is total thus does not suffice to ensure that the set $\text{opt}_< (A \times B)$ of OT optimal candidate pairs in $A \times B$ is also a singleton. The commutativity identity (13) could thus in principle fail, because its right-hand side $\text{opt}_< A \times \text{opt}_< B$ is a singleton while its left-hand side $\text{opt}_< (A \times B)$ is not.

But Prince's Proposition 1 ensures that can actually never happen: the two different candidate pairs (α, β) and $(\widehat{\alpha}, \widehat{\beta})$ which share the same summed tuple $\mathbf{a} + \mathbf{b} = \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$ of constraint violations can never belong to the set $\text{opt}_< (A \times B)$ of OT optimal candidate pairs in $A \times B$. In fact, let us assume by contradiction that they do. Without loss of generality, we assume that the two candidates α and $\widehat{\alpha}$ are different (analogous considerations hold if it is the two candidates β and $\widehat{\beta}$ that are different instead). Since the candidate set A is distinctive, the tuples $\mathbf{a}$ and $\widehat{\mathbf{a}}$ of constraint violations of the two candidates α and $\widehat{\alpha}$ must be different. Since $<$ is total, one of these two tuples is larger than the other relative to $<$. Without loss of generality, we suppose that $\mathbf{a} < \widehat{\mathbf{a}}$. Crucially, this assumption $\mathbf{a} < \widehat{\mathbf{a}}$ entails that $\mathbf{a} + \widehat{\mathbf{b}} < \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$, by reasoning as above from (14) to (16). In other words, the candidate pair $(\alpha, \widehat{\beta})$ beats the candidate pair $(\widehat{\alpha}, \widehat{\beta})$. This conclusion contradicts the assumption that the candidate pair $(\widehat{\alpha}, \widehat{\beta})$ is OT optimal in $A \times B$.

5 CONSTRAINT SUMMATION IS TYPOLOGICALLY INNOCUOUS: BEYOND OPTIMALITY THEORY

In Section 3, we have formalized typological innocuousness of the constraint summation assumption (1) used by DT and the OPM through the commutativity identity (13). In Section 4, we have recalled from Prince (2015) that this identity holds in the case of the OT model of constraint interaction. In other words, despite constraint summation, theories such as DT and the OPM make the same typological predictions as classical OT when the constraint set only consists of classical constraints and no distinctiveness or OP faithfulness constraints, whereby constraint summation is typologically innocuous.

The focus on OT so far was motivated by the fact that it is the most widely adopted version of constraint-based phonology, and indeed the one adopted in Flemming's implementation of DT and in McCarthy's implementation of the OPM. Yet, the more recent constraint-based phonological literature (Pater 2009; Potts *et al.* 2010) has advocated variants of OT where optimum selection is based on linear utility functions, as foreshadowed in Goldsmith (1990, §6.5) and Goldsmith (1991, page 259) and advocated in *Linear OT* (LOT; Keller 2000, 2006) and *Harmonic Grammar* (HG; Legendre *et al.* 1990b,a; Smolensky and Legendre 2006). Does the typological innocuousness of the constraint summation assumption (1) extend beyond OT to these alternative implementations of constraint-based phonology? In other words, is the commutativity identity (13) specific to OT's lexicographic order or does it extend to other ways of ordering tuples of constraint violations? This section addresses this question.

Here is a preview of the core result. In Subsection 4.2, we have used two properties of the lexicographic order to establish the commutativity identity (13) for OT. The first property is that the lexicographic ordering of two tuples of constraint violations is not affected by adding the same quantities to the constraint violations in the two tuples, whereby the inequalities (14) entail those in (16). Subsection 5.1 generalizes this property into the notion of *additive* orders. The second property of the lexicographic order that we have used in Subsection 4.2 to establish the commutativity identity for OT is that it is total. This means that any two tuples of constraint violations which are not ordered (neither is larger than the other) must be identical. Subsection 5.2 generalizes total orders to *weak* orders: tuples which are not ordered need not be identical but must be *equivalent*, namely need to satisfy some generalization of the notion of identity. Subsection 5.4 finally shows that additive weak orders are the minimal structure required by a constraint-based phonological formalism to satisfy the commutativity identity (13) and thus to ensure the typological innocuousness of the constraint summation assumption made by DT and the OPM. The proof of this result relies on some properties of additive weak orders established in Subsection 5.3.

Additive orders 5.1

Throughout this section, we consider an arbitrary *strict order* $<$. This means that $<$ satisfies the following three conditions for any tuples $\mathbf{a}, \mathbf{b}, \mathbf{c}$ of constraint violations: it is *irreflexive*, namely $\mathbf{a} < \mathbf{a}$ never holds; it is *asymmetric*, namely $\mathbf{a} < \mathbf{b}$ and $\mathbf{b} < \mathbf{a}$ never both hold; it is *transitive*, namely $\mathbf{a} < \mathbf{b}$ and $\mathbf{b} < \mathbf{c}$ entail $\mathbf{a} < \mathbf{c}$. Recall from (10) that $\mathbf{a} + \mathbf{b} = (a_1 + b_1, \ldots, a_n + b_n)$ denotes the component-wise sum of two tuples $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ of constraint violations. The implication (17) captures the intuitive idea that, if $\mathbf{a}$ is smaller than $\mathbf{b}$ and if the same quantity $\mathbf{c}$ is added to both, the resulting sum $\mathbf{a} + \mathbf{c}$ ought to be smaller than the sum $\mathbf{b} + \mathbf{c}$. Although intuitive, it is possible to construct orders which fail at this condition (one such example is provided in Subsection 6.3). A strict order $<$ which satisfies condition (17) for any three tuples $\mathbf{a}, \mathbf{b}, \mathbf{c}$ of constraint violations is called *additive* (Anderson and Feil 1988).

(17) If: $\mathbf{a} < \mathbf{b}$,
then: $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{c}$.

To illustrate, OT's lexicographic order is additive (see Subsection 6.1 below for more details).

Throughout this section, we assume that constraint violations can be either positive or negative integers. In other words, we assume that the order $<$ is defined over arbitrary tuples of integers, not necessarily nonnegative integers.[9] This assumption effectively means that in the consequent of the additivity condition (17), we can either add to or subtract from the constraint violations listed in the tuples $\mathbf{a}$ and $\mathbf{b}$. This flexibility will be crucial for some of the reasoning developed in this section, such as the proof of condition (21) below. This assumption does not restrict the scope of our results, because the orders of interest considered in Subsection 6 (such as OT's lexicographic order and HG's order based on linear utility functions) can indeed all be construed as ranging over tuples of positive and negative numbers.

The additivity condition (17) entails the variant in (18) for any four tuples $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ of constraint violations.

(18) If: $\mathbf{a} < \mathbf{b}$ and $\mathbf{c} < \mathbf{d}$,
then: $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{d}$.

In fact, the assumption $\mathbf{a} < \mathbf{b}$ in the antecedent of (18) ensures that $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{c}$ through the additivity condition (17). Analogously, the assumption $\mathbf{c} < \mathbf{d}$ ensures that $\mathbf{b}+\mathbf{c} < \mathbf{b}+\mathbf{d}$. Finally, the two conditions $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{c}$ and $\mathbf{b} + \mathbf{c} < \mathbf{b} + \mathbf{d}$ thus obtained ensure the consequent $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{d}$ of (18), because the order $<$ is transitive.

[9] The additivity assumption (17) thus means that $(\mathbb{Z}^n, +, <)$ is an *ordered group* (Anderson and Feil 1988).

Weak orders 5.2

As motivated in Subsection 3.1 (see in particular footnote 6), we allow for the possibility that the strict order $<$ is *partial*, not necessarily total. This means that there can exist two tuples $\mathbf{a}, \mathbf{b}$ of constraint violations such that neither $\mathbf{a} < \mathbf{b}$ nor $\mathbf{b} < \mathbf{a}$. In this case, we say that $\mathbf{a}$ and $\mathbf{b}$ are *incommensurable* (relative to $<$) and we write $\mathbf{a} \sim \mathbf{b}$. In other words, the partial strict order defines a corresponding *incommensurability relation* $\sim$, as in (19).

(19) $\mathbf{a} \sim \mathbf{b}$ if and only if neither $\mathbf{a} < \mathbf{b}$ nor $\mathbf{b} < \mathbf{a}$.

Since the strict order $<$ is irreflexive, the inequality $\mathbf{a} < \mathbf{a}$ fails for any tuple $\mathbf{a}$ of constraint violations. In other words, any tuple $\mathbf{a}$ is incommensurable with itself and the incommensurability relation $\sim$ is therefore reflexive. Furthermore, the incommensurability relation $\sim$ is obviously symmetric. The strict order $<$ is called *weak* provided the corresponding incommensurability relation $\sim$ is also transitive, namely qualifies as an equivalence relation among tuples of constraint violations (Roberts and Tesman 2005, section 4.2.4). We will see some examples below in Subsection 6.

The intuition behind this definition is that a weak order $<$ orders two incommensurable tuples in the same way relative to any other tuples, in the sense that the implication (20) holds for any three tuples $\mathbf{a}, \mathbf{b}, \mathbf{c}$ of constraint violations.

(20) If: $\mathbf{a} < \mathbf{b}$ and $\mathbf{b} \sim \mathbf{c}$,
then: $\mathbf{a} < \mathbf{c}$.

In fact, let us assume by contradiction that the consequent $\mathbf{a} < \mathbf{c}$ of (20) fails. This means that either $\mathbf{a} \sim \mathbf{c}$ or $\mathbf{c} < \mathbf{a}$. But $\mathbf{a} \sim \mathbf{c}$ is impossible, because together with the assumption $\mathbf{b} \sim \mathbf{c}$ and the transitivity of $\sim$, it would entail $\mathbf{a} \sim \mathbf{b}$, contradicting the other assumption $\mathbf{a} < \mathbf{b}$. Analogously, $\mathbf{c} < \mathbf{a}$ is impossible as well, because together with the assumption $\mathbf{a} < \mathbf{b}$ and the transitivity of $<$, it would entail $\mathbf{c} < \mathbf{b}$, contradicting the other assumption $\mathbf{b} \sim \mathbf{c}$.

A characterization of additive weak orders 5.3

We now have two assumptions on the strict partial order $<$ over tuples of (positive and negative) constraint violations: that it is additive and that it is a weak order. Subsection 5.4 will show that these two assumptions are necessary and sufficient to guarantee the commutativity identity (13) and thus to ensure that the constraint summation assumption (1) in DT and the OPM is typologically innocuous. Towards establishing this result, we now take a closer look at the combination of these two assumptions that an order is both additive and weak.

Suppose that the strict order $<$ is additive, in the sense that it satisfies condition (17). Its corresponding incommensurability relation $\sim$ is then additive as well, in the sense that it satisfies the completely analogous condition (21) for any tuples $\mathbf{a}, \mathbf{b}, \mathbf{c}$ of constraint violations.

(21) If: $\mathbf{a} \sim \mathbf{b}$,
then: $\mathbf{a} + \mathbf{c} \sim \mathbf{b} + \mathbf{c}$.

In fact, let us assume by contradiction that the consequent $\mathbf{a}+\mathbf{c} \sim \mathbf{b}+\mathbf{c}$ of (21) fails. This means that either $\mathbf{b} + \mathbf{c} < \mathbf{a} + \mathbf{c}$ or $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{c}$. For concreteness, let us suppose that the former case $\mathbf{b} + \mathbf{c} < \mathbf{a} + \mathbf{c}$ holds. Adding $-\mathbf{c}$ to both sides (which we are allowed to do, because we are not resrticting ourselves to nonnegative constraint violations) yields $\mathbf{b} < \mathbf{a}$, because the order $<$ satisfies the additivity condition (17). This conclusion $\mathbf{b} < \mathbf{a}$ contradicts the assumption $\mathbf{a} \sim \mathbf{b}$.

The reasoning used in Subsection 5.1 to show that the original additivity condition (17) for the order $<$ entails the variant (18) can be rebooted here to show that the additivity condition (21) for the incommensurability relation $\sim$ entails the analogous variant (22) for any four tuples $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ of constraint violations.

(22) If: $\mathbf{a} \sim \mathbf{b}$ and $\mathbf{c} \sim \mathbf{d}$,
then: $\mathbf{a} + \mathbf{c} \sim \mathbf{b} + \mathbf{d}$.

In fact, the assumption $\mathbf{a} \sim \mathbf{b}$ in the antecedent of (22) ensures that $\mathbf{a} + \mathbf{c} \sim \mathbf{b} + \mathbf{c}$ through the additivity condition (21). Analogously, the assumption $\mathbf{c} \sim \mathbf{d}$ ensures that $\mathbf{b}+\mathbf{c} \sim \mathbf{b}+\mathbf{d}$. Finally, the two conditions $\mathbf{a} + \mathbf{c} \sim \mathbf{b} + \mathbf{c}$ and $\mathbf{b} + \mathbf{c} \sim \mathbf{b} + \mathbf{d}$ thus obtained ensure the consequent $\mathbf{a} + \mathbf{c} \sim \mathbf{b} + \mathbf{d}$ of (22), because the incommensurability relation $\sim$ is transitive.

Conditions (17)/(18) and (21)/(22) characterize additivity of the order $<$ and of the incommensurability relation $\sim$ *separately*. They entail the following mixed additivity condition (23), which features the two relations jointly. This condition says that the validity of an inequality is not affected by adding incommensurable elements at both sides.

(23) If: $\mathbf{a} < \mathbf{b}$ and $\mathbf{c} \sim \mathbf{d}$,
then: $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{d}$.

In fact, the assumption $\mathbf{a} < \mathbf{b}$ in the antecedent of (23) ensures that $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{c}$ through the additivity condition (17) for the order $<$. Analogously, the assumption $\mathbf{c} \sim \mathbf{d}$ ensures that $\mathbf{b}+\mathbf{c} \sim \mathbf{b}+\mathbf{d}$, through the additivity condition (21) for the incommensurability relation $\sim$. Finally, the two conditions $\mathbf{a}+\mathbf{c} < \mathbf{b}+\mathbf{c}$ and $\mathbf{b}+\mathbf{c} \sim \mathbf{b}+\mathbf{d}$ thus obtained ensure the consequent $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{d}$ of (23), because of the condition (20) that incommensurable tuples are ordered alike.

As noted above, since the strict order $<$ is irreflexive, its incommensurability relation $\sim$ is reflexive. This means in particular that $\mathbf{c} \sim \mathbf{d}$ whenever $\mathbf{c} = \mathbf{d}$. The mixed additivity condition (23) thus generalizes the original additivity condition (17) for the weak order $<$ from the special case $\mathbf{c} = \mathbf{d}$ to the more general case $\mathbf{c} \sim \mathbf{d}$. This generalization makes intuitive sense because the assumption that $<$ is a weak order means that its incommensurability relation $\sim$ is an equivalence relation, namely that $\sim$ generalizes the identity $=$.

We conclude this subsection with the characterization of additive weak orders provided by the following lemma, in terms of the two additivity conditions (17) and (21) or equivalently in terms of the mixed additivity condition (23). The next subsection will use this characterization to establish a connection between additive weak orders and the commutativity identity (13) which was shown to be crucial for the typological innocuousness of the constraint summation assumption made by theories such as DT and the OPM.

LEMMA 1 *A strict (possibly partial) order $<$ is an additive weak order if and only if it satisfies the additivity condition (17) and furthermore its incommensurability relation satisfies the additivity condition (21). Equivalently, if and only if it satisfies the mixed additivity condition (23).*

PROOF We only need to show that the mixed additivity condition (23) entails transitivity of the incommensurability relation $\sim$. Let us assume by contradiction that $\sim$ is not transitive, namely that $\mathbf{a} \sim \mathbf{b}$ and $\mathbf{b} \sim \mathbf{c}$ but $\mathbf{a} \not\sim \mathbf{c}$. The latter condition $\mathbf{a} \not\sim \mathbf{c}$ means that either

$\mathbf{a} < \mathbf{c}$ or $\mathbf{c} < \mathbf{a}$; for concreteness, we assume that the former case holds. By (23), $\mathbf{a} < \mathbf{c}$ and $\mathbf{c} \sim \mathbf{b}$ entail $\mathbf{a} + \mathbf{c} < \mathbf{c} + \mathbf{b}$. By (23) again, $\mathbf{a}+\mathbf{c} < \mathbf{c}+\mathbf{b}$ and $-\mathbf{c} \sim -\mathbf{c}$ (which holds because $\sim$ is reflexive, as it is the incommensurability relation of a strict order), entail $\mathbf{a} < \mathbf{b}$. This conclusion $\mathbf{a} < \mathbf{b}$ contradicts the hypothesis $\mathbf{a} \sim \mathbf{b}$. □

5.4 *The commutativity identity (13) holds for (and only for) additive weak orders*

This section proves the following Proposition 2, which is the main result of this paper. The "if" statement of the proposition says that additive weak orders provide *sufficient* structure to ensure the commutativity identity (13). Furthermore, the "only if" statement says that additive weak orders provide the *necessary* structure for the commutativity identity (13) to hold. In other words, the constraint summation assumption (1) made by DT and the OPM is typologically innocuous if and only if tuples of constraint violations are compared and optimized relative to an additive weak order.

PROPOSITION 2 *Consider a strict (possibly partial) order $<$ over tuples of constraint violations. The commutativity identity (13) repeated below holds for any two candidate sets A and B if and only if $<$ is an additive weak order.*

(13) $$\boxed{\operatorname*{opt}_{<}(A \times B) = \operatorname*{opt}_{<} A \times \operatorname*{opt}_{<} B}$$

PROOF In order to streamline the proof of the proposition, it useful to split the commutativity identity (13) into the two inclusions (24).

(24) a. $\operatorname*{opt}_{<}(A \times B) \subseteq \operatorname*{opt}_{<} A \times \operatorname*{opt}_{<} B$

b. $\operatorname*{opt}_{<}(A \times B) \supseteq \operatorname*{opt}_{<} A \times \operatorname*{opt}_{<} B$

The proof of the proposition relies on the characterization of additive weak orders provided by Lemma 1 through the additivity condition (17) for the order $<$ and the additivity condition (21) for the incommensurability relation $\sim$, as summarized in (25). The additivity condition (17) for $<$ suffices to derive the inclusion (24a) while both additivity conditions (17) and (21) for $<$ and $\sim$ are needed to derive the reverse inclusion (24b). Vice versa, the inclusions (24a) and (24b) each suffice to derive the additivity conditions (17) and (21) for $<$ and $\sim$, respectively.

(25) a. $<$ additivity condition (17) $\Longrightarrow$ inclusion (24a)

b. $\left.\begin{array}{l} < \text{additivity condition (17)} \\ \sim \text{additivity condition (21)} \end{array}\right\}$ $\Longrightarrow$ inclusion (24b)

c. $<$ additivity condition (17) $\Longleftarrow$ 4 inclusion (24a)

d. $\sim$ additivity condition (21) $\Longleftarrow$ inclusion (24b)

We start by showing that the additivity condition (17) for $<$ entails the inclusion (24a), as stated in (25a). We consider a candidate pair $(\widehat{\alpha}, \widehat{\beta})$ that belongs to the set $\text{opt}_<(A \times B)$ of optimal candidate pairs of $A \times B$. We suppose by contradiction that either the candidate $\widehat{\alpha}$ does not belong to the set $\text{opt}_< A$ of optimal candidates of A or the candidate $\widehat{\beta}$ does not belong to the set $\text{opt}_< B$ of optimal candidates of B (or both). For concreteness, we assume that the former case holds, namely that $\widehat{\alpha}$ does not belong to the optimal set $\text{opt}_< A$. This means in turn that there exists another candidate α of A such that $4\mathrm{a} < \mathbf{a}$, where $\mathbf{a}$ and $\widehat{\mathbf{a}}$ are the tuples of constraint violations of the two candidates α and $\widehat{\alpha}$, respectively. Since $<$ satisfies the additivity condition (17), $\mathbf{a} < \widehat{\mathbf{a}}$ entails $\mathbf{a} + \widehat{\mathbf{b}} < \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$, where $\widehat{\mathbf{b}}$ is the tuple of constraint violations of the candidate $\widehat{\beta}$. This inequality $\mathbf{a} + \widehat{\mathbf{b}} < \widehat{\mathbf{a}} + \widehat{\mathbf{b}}$ says that the candidate pair $(\alpha, \widehat{\beta})$ beats the candidate pair $(\widehat{\alpha}, \widehat{\beta})$. This conclusion contradicts the hypothesis that the candidate pair $(\widehat{\alpha}, \widehat{\beta})$ belongs to the set $\text{opt}_<(A \times B)$ of optimal candidate pairs of $A \times B$. This reasoning is analogous to the reasoning used in Subsection 4.1 to prove the inclusion (15a).

We show next that the two additivity conditions (17) and (21) – and their corollaries (18), (22), and (23 – entail the other inclusion (24b), as stated in (25b). Consider a candidate $\widehat{\alpha}$ that belongs to the set $\text{opt}_< A$ of optimal candidates of A. Consider next a candidate $\widehat{\beta}$ that belongs to the set $\text{opt}_< B$ of optimal candidates of B. We assume by contradiction that the candidate pair $(\widehat{\alpha}, \widehat{\beta})$ does not belong to the set $\text{opt}_<(A \times B)$ of optimal candidate pairs of $A \times B$. This means that there exists another candidate pair (α, β) in $A \times B$ such that the sum $\mathbf{a} + \mathbf{b}$ of the tuples $\mathbf{a}$ and $\mathbf{b}$ of constraint violations of candidates α and β is smaller than the sum $\widehat{\mathbf{a}} + \widehat{\mathbf{b}}$ of the tuples $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{b}}$ of constraint violations of the candidates $\widehat{\alpha}$ and $\widehat{\beta}$, namely $\mathbf{a}+\mathbf{b} < \widehat{\mathbf{a}}+\widehat{\mathbf{b}}$. Since the candidate α belongs to A and the candidate $\widehat{\alpha}$ is optimal for A, either $\widehat{\mathbf{a}} < \mathbf{a}$ or else $\mathbf{a} \sim \widehat{\mathbf{a}}$. Analogously, since the candidate β belongs to B and the candidate $\widehat{\beta}$ is optimal for B, either $\widehat{\mathbf{b}} < \mathbf{b}$ or else $\mathbf{b} \sim \widehat{\mathbf{b}}$. If $\widehat{\mathbf{a}} < \mathbf{a}$ and $\widehat{\mathbf{b}} < \mathbf{b}$, the additivity condition (18) entails $\widehat{\mathbf{a}}+\widehat{\mathbf{b}} < \mathbf{a}+\mathbf{b}$, which contradicts the assumption $\mathbf{a}+\mathbf{b} < \widehat{\mathbf{a}}+\widehat{\mathbf{b}}$. If $\widehat{\mathbf{a}} < \mathbf{a}$ and $\widehat{\mathbf{b}} \sim \mathbf{b}$ (or if $\widehat{\mathbf{a}} \sim \mathbf{a}$ and $\widehat{\mathbf{b}} < \mathbf{b}$), the mixed additivity condition (23) entails $\widehat{\mathbf{a}}+\widehat{\mathbf{b}} < \mathbf{a}+\mathbf{b}$, which again contradicts the assumption $\mathbf{a}+\mathbf{b} < \widehat{\mathbf{a}}+\widehat{\mathbf{b}}$. Finally, if $\widehat{\mathbf{a}} \sim \mathbf{a}$ and $\widehat{\mathbf{b}} \sim \mathbf{b}$, the additivity condition (22) for the incommensurability relation entails $\widehat{\mathbf{a}}+\widehat{\mathbf{b}} \sim \mathbf{a}+\mathbf{b}$, which again contradicts the assumption $\mathbf{a}+\mathbf{b} < \widehat{\mathbf{a}}+\widehat{\mathbf{b}}$. This reasoning is analogous to the reasoning used in Subsection 4.1 to prove the inclusion (15b).

Turning to the opposite direction, we show now that the inclusion (24a) entails that the order $<$ satisfies the additivity condition (17), as stated in (25c). Thus, we assume that the antecedent $\mathbf{a} < \mathbf{b}$ of the additivity condition (17) holds and we consider an arbitrary third vector $\mathbf{c}$. We consider a set $A = \{\alpha, \beta\}$ consisting of two candidates α and β whose tuples of constraint violations are $\mathbf{a}$ and $\mathbf{b}$. Furthermore, we consider a set $B = \{\gamma\}$ consisting of a unique candidate γ whose tuple of constraint violations is $\mathbf{c}$. The hypothesis $\mathbf{a} < \mathbf{b}$ means that the set $\text{opt}_< A$ of optimal candidates of A only consists of the candidate α. Furthermore, the set $\text{opt}_< B$ of optimal candidates of B only consists of the candidate γ, because B is a singleton. Hence, the product $\text{opt}_< A \times \text{opt}_< B$ of the two optimal sets only consists of the pair (α, γ). Finally, $A \times B = \{(\alpha, \gamma), (\beta, \gamma)\}$. The inclusion (24a) thus says that the set $\text{opt}_< (A \times B)$ of optimal candidate pairs of $A \times B$ only consists of the pair (α, γ) and does not contain the other pair (β, γ). This means in turn that neither $\mathbf{a}+\mathbf{c} \sim \mathbf{b}+\mathbf{c}$ nor $\mathbf{b}+\mathbf{c} < \mathbf{a}+\mathbf{c}$ and thus that $\mathbf{a}+\mathbf{c} < \mathbf{b}+\mathbf{c}$. This conclusion shows that the additivity condition (17) holds, namely that $\mathbf{a} < \mathbf{b}$ entails $\mathbf{a}+\mathbf{c} < \mathbf{b}+\mathbf{c}$.

We conclude by showing that the inclusion (24b) entails that the incommensurability relation $\sim$ satisfies the additivity condition (21), as stated in (25d). Thus, we assume that the antecedent $\mathbf{a} \sim \mathbf{b}$ of the additivity condition (21) holds and we consider an arbitrary third vector $\mathbf{c}$. We consider the candidate sets $A = \{\alpha, \beta\}$ and $B = \{\gamma\}$ as above, where the three candidates α, β, γ have the tuples of constraint violations $\mathbf{a}, \mathbf{b}, \mathbf{c}$, respectively. The incommensurability assumption $\mathbf{a} \sim \mathbf{b}$

says that $\text{opt}_< A = A$. Furthermore, $\text{opt}_< B = B$, because B is a singleton. Hence, $\text{opt}_< A \times \text{opt}_< B = A \times B$. The inclusion (24b) thus says that $A \times B = \text{opt}_< (A \times B)$. In other words, both candidate pairs (α, β) and (α, γ) of the set $A \times B$ actually belong to the optimal set $\text{opt}_< (A \times B)$. This means in turn that the tuples of constraint violations of these two candidate pairs (α, β) and (α, γ) are incommensurable, namely that $\mathbf{a} + \mathbf{c} \sim \mathbf{b} + \mathbf{c}$. This conclusion shows that the additivity condition (21) holds, namely that $\mathbf{a} \sim \mathbf{b}$ entails $\mathbf{a} + \mathbf{c} \sim \mathbf{b} + \mathbf{c}$. □

APPLICATIONS 6

This section re-derives Prince's proposition 1 that OT satisfies the commutativity identity (13) as a special case of Proposition 2 obtained in the preceding section. Furthermore, it shows that the commutativity identity extends to constraint-based phonological frameworks that order tuples of constraint violations based on additive utility functions. It follows in particular (see Proposition 3) that the commutativity identity holds for HG. In other words, the typological innocuousness of the constraint summation assumption made by DT and the OPM extends from the OT to the HG mode of constraint interaction.

Re-deriving Prince's result for OT 6.1

Any strict order which is *total* (namely defined for any pair of different tuples of constraint violations) is in particular a weak order. In fact, totality means that two tuples of constraint violations are incommensurable only if they are identical, whereby the incommensurability relation $\sim$ coincides with the identity and it is therefore transitive. Proposition 2 thus ensures that the commutativity identity (13) crucial for DT and the OPM holds whenever grammatical optimization is relative to a total additive strict order.

As our first application of Proposition 2, we can now derive anew Prince's Proposition 1 for OT. In fact, let $<$ be OT's lexicographic order corresponding to some ranking of the n constraints. We assume without loss of generality that C_1 is ranked at the top, followed by C_2, and so on. As reviewed in Subsection 4.1, the condition $\mathbf{a} < \mathbf{b}$ then holds for two tuples $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ of constraint violations if and only if conditions (26) hold for some $k \in \{1, \ldots, n\}$.

(26)
$$\begin{array}{rcl} a_1 & = & b_1 \\ & \vdots & \\ a_k & = & b_k \\ a_{k+1} & < & b_{k+1} \end{array}$$

The lexicographic order $<$ is total, namely defined for any two different tuples of constraint violations. Furthermore, it is additive, namely it satisfies the implication (17): the assumption $\mathbf{a} < \mathbf{b}$ that (26) holds entails the conclusion $\mathbf{a} + \mathbf{c} < \mathbf{b} + \mathbf{c}$ that (27) holds as well.

(27)
$$\begin{array}{rcl} a_1 + c_1 & = & b_1 + c_1 \\ & \vdots & \\ a_k + c_k & = & b_k + c_k \\ a_{k+1} + c_{k+1} & < & b_{k+1} + c_{k+1} \end{array}$$

Prince's Proposition 1 for OT thus follows as a special case of Proposition 2: the commutativity identity (13) holds in the case of OT because grammatical optimization in OT is computed relative to the lexicographic order which is additive and total.

6.2 *Extension to HG*

To explore further applications of Proposition 2, we consider a *utility function U* which assigns to each tuple $\mathbf{a}$ of constraint violations a number $U(\mathbf{a})$. We can then order the tuples of constraint violations based on their utility, with smaller tuples corresponding to a smaller utility, as in (28).

(28) $\mathbf{a} < \mathbf{b}$ if and only if $U(\mathbf{a}) < U(\mathbf{b})$.

The resulting relation $<$ is obviously a strict order. It is partial, because tuples of constraint violations which achieve the same utility are incommensurable. Furthermore, it is a weak order, because the corresponding incommensurability relation $\sim$ described in (29) is obviously transitive.

(29) $\mathbf{a} \sim \mathbf{b}$ if and only if $U(\mathbf{a}) = U(\mathbf{b})$.

Suppose that the utility function U is *additive*, namely that the identity $U(\mathbf{a}+\mathbf{b}) = U(\mathbf{a}) + U(\mathbf{b})$ holds for any tuples $\mathbf{a}, \mathbf{b}$ of constraint violations. In this case, the corresponding weak strict order $<$ satisfies the additivity condition (17). In fact, the assumption $\mathbf{a} < \mathbf{b}$ of this additivity condition means that $U(\mathbf{a}) < U(\mathbf{b})$. Hence $U(\mathbf{a}) + U(\mathbf{c}) < U(\mathbf{b}) + U(\mathbf{c})$. By additivity, this means $U(\mathbf{a}+\mathbf{c}) < U(\mathbf{b}+\mathbf{c})$, whereby $\mathbf{a}+\mathbf{c} < \mathbf{b}+\mathbf{c}$. Proposition 2 thus ensures that the commutativity identity (13) crucial for DT and the OPM holds whenever grammatical optimization is relative to the order induced by an additive utility function.

Taking advantage of the fact that constraint violations are integers, Magri (2020) shows (through a simple twist of the Fundamental Theorem of Linear Algebra; Strang 2006, Section 2.6) that for any additive utility function U, there exist a *weight vector* $\mathbf{w} = (w_1, \ldots, w_n)$ such that the utility $U(\mathbf{a})$ of any tuple of integer constraint violations can be described as the weighted sum of the constraint violations collected in the tuple $\mathbf{a}$, namely $U(\mathbf{a}) = \sum_{i=1}^{n} a_i w_i$. In other words, the partial strict order corresponding to an additive utility function in the sense of (28) yields the HG model of grammatical optimization (Legendre *et al.* 1990b,a; Smolensky and Legendre 2006; Pater 2009; Potts *et al.* 2010). Proposition 2 thus ensures that the commutativity identity (13) crucial for DT and the OPM extends from OT to HG, as stated by the following proposition.

PROPOSITION 3 *The commutativity identity (13) holds for any two candidate sets A and B relative to HG's order $<$ corresponding to any constraint weighting: a candidate $\widehat{\alpha}$ belongs to the set* $\mathrm{opt}_< A$ *of optimal candidates of A relative to the HG order $<$ corresponding to that weighting and a candidate $\widehat{\beta}$ belongs to the set* $\mathrm{opt}_< B$ *of optimal candidates of B if and only if the candidate pair $(\widehat{\alpha}, \widehat{\beta})$ belongs to the set* $\mathrm{opt}_< (A \times B)$ *of optimal candidate pairs in $A \times B$, when candidate pairs are compared based on summed constraint violations.*

When the commutativity identity fails 6.3

Crucially, Proposition 2 provides not only a sufficient but also a necessary condition for the commutativity identity (13) to hold. Thus, this proposition can be used not only to verify that the commutativity identity holds, as we have done so far, but also to disprove that it does. To illustrate, suppose that there are only $n = 2$ constraints and consider the *quadratic* utility function U defined as in (30) for any pair $\mathbf{a} = (a_1, a_2)$ of constraint violations.

(30) $U(\mathbf{a}) = a_1^2 + a_2^2$

The corresponding relation $<$ as in (28) is a weak partial strict order. Yet, it does not satisfy the additivity condition (17). In fact, consider for instance $\mathbf{a} = (2, 2)$, $\mathbf{b} = (0, 3)$ and $\mathbf{c} = (4, 4)$. In this case, $\mathbf{a} < \mathbf{b}$ (because $U(\mathbf{a}) = 2^2+2^2 = 8$ while $U(\mathbf{b}) = 0^2+3^2 = 9$). But $\mathbf{b}+\mathbf{c} < \mathbf{a}+\mathbf{c}$ (because $U(\mathbf{a}+\mathbf{c}) = 6^2+6^2 = 72$ while $U(\mathbf{b}+\mathbf{c}) = 4^2+7^2 = 65$). Proposition 2 therefore ensures that the commutativity identity (13) crucial for DT and the OPM fails when constraint violations are optimized relative to the order induced by the quadratic utility function (30).

7 CONCLUSIONS

Usually in the constraint-based phonological literature (starting with Prince and Smolensky 1993/2004), each underlying form comes with a preassigned set of candidate surface realizations. Each of these candidates is represented as a tuple of constraint violations. These tuples are compared according to some strict (possibly partial) order $<$ that extends the notion of "being smaller than" from single numbers to tuples of numbers. The optimal candidate for a given underlying form is the one which violates the constraints the least, namely the one with the smallest tuple of constraint violations.

DT (Flemming 2002, 2004, 2008) enriches the classical constraint set with distinctiveness constraints. Furthermore, approaches to paradigm uniformity effects such as the OPM (Kenstowicz 1997; McCarthy 2005) enrich the classical constraint set with OP faithfulness constraints. Crucially, classical (faithfulness and markedness) constraints evaluate a single candidate surface form at a time while distinctiveness and OP faithfulness constraints evaluate multiple candidate surface forms simultaneously, relative to their contrastiveness and their similarity, respectively. As a consequence, the classical constraints need to be "lifted" from a single candidate to multiple candidates. A reasonable way to do that is to sum their violations across multiple candidates.

Does this assumption of constraint summation made by DT and the OPM make sense? We have formulated this question as follows. Suppose that we restrict ourselves to a constraint set which includes no distinctiveness or OP faithfulness constraints but only classical (markedness and faithfulness) constraints. In this case, can we guarantee that the typological predictions of DT and the OPM coincide with those of the classical theory, despite constraint summation? In other words, is the assumption of constraint summation made by DT and the OPM typologically innocuous?

This paper has shown that constraint summation is indeed typologically innocuous if and only if constraint optimization is performed relative to an order $<$ of tuples of constraint violations which is additive and weak. In other words, additive weak orders provide the "minimal structure" (to go back to Talagrand's admonition in the quote at the beginning of the paper) for typological innocuousness to hold. This technical condition on grammatical optimization is verified for instance in the case of OT and HG. Our result extends and systematizes an earlier independent result for OT obtained by Prince (2015). Our result provides a solid foundation for theories such as DT and the OPM which make use of constraint summation, for a large class of modes of constraint interaction.

REFERENCES

Adam ALBRIGHT (2010), Base-driven leveling in Yiddish verb paradigms, *Natural Language & Linguistic Theory*, 28(3):475–537.

Marlow ANDERSON and Todd FEIL (1988), *Lattice ordered groups: an introduction*, D. Reidel Publishing Company, Dordrecht.

Laure BENUA (1997), *Transderivational identity: phonological relations between words*, Ph.D. thesis, University of Massachusetts, Amherst.

Juliette BLEVINS (2004), *Evolutionary phonology: the emergence of sound patterns*, Cambridge University Press, Cambridge.

Paul BOERSMA and Silke HAMANN (2008), The evolution of auditory dispersion in bidirectional constraint grammars, *Phonology*, 25:217–270.

Jon P. DAYLEY (1989), *Tümpisa (Panamint) Shoshone grammar*, University of California Press, Berkley.

Edward FLEMMING (2002), *Auditory representations in phonology*, Routledge.

Edward FLEMMING (2004), Contrast and perceptual distinctiveness, in Bruce HAYES, Robert KIRCHNER, and Donca STERIADE, editors, *Phonetically-based phonology*, pp. 232–276, Cambridge University Press, Cambridge.

Edward FLEMMING (2008), The realized input, unpublished manuscript, MIT.

Edward FLEMMING (2017a), Dispersion effects in phonology, `https://sites.google.com/site/parisseminarcalt/home/edwardscourse`, handouts for a three-day course in Paris.

Edward FLEMMING (2017b), Dispersion Theory and phonology, in Mark ARONOFF, editor, *The Oxford research encyclopedia of linguistics*, Oxford University Press, Oxford.

John A. GOLDSMITH (1990), *Autosegmental and metrical phonology*, Basil Blackwell, Oxford.

John A. GOLDSMITH (1991), Phonology as an intelligent system, in Donna Jo NAPOLI and Judy KEGL, editors, *Bridges between psychology and linguistics: a Swarthmore festschrift for Lila Gleitman*, pp. 247–267, Lawrence Erlbaum, Mahwah, NJ.

Gregory K. IVERSON and Joseph C. SALMONS (1996), Mixtec prenasalization as hyper-voicing, *International Journal of American Linguistics*, 62:165–175.

Frank KELLER (2000), *Gradience in Grammar. Experimental and computational aspects of degrees of grammaticality*, Ph.D. thesis, University of Edinburgh.

Frank KELLER (2006), Linear Optimality Theory as a model of gradience in grammar, in Gisbert FANSELOW, Caroline FÉRY, Ralph VOGEL, and Matthias SCHLESEWSKY, editors, *Gradience in grammar: generative perspectives*, pp. 270–287, Oxford University Press, Oxford.

Michael KENSTOWICZ (1997), Base identity and uniform exponence: alternatives to cyclicity, in Jacques DURAND and Bernard LAKS, editors, *Current trends in phonology: models and methods*, pp. 363–394, Salford: University of Salford.

Géraldine LEGENDRE, Yoshiro MIYATA, and Paul SMOLENSKY (1990a), Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: an application, in Morton Ann GERNSBACHER and Sharon J. DERRY, editors, *Proceedings of the 12th annual conference of the Cognitive Science Society*, pp. 884–891, Lawrence Erlbaum Associates, Hillsdale, NJ.

Géraldine LEGENDRE, Yoshiro MIYATA, and Paul SMOLENSKY (1990b), Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: theoretical foundations, in Morton Ann GERNSBACHER and Sharon J. DERRY, editors, *Proceedings of the 12th annual conference of the Cognitive Science Society*, pp. 388–395, Lawrence Erlbaum, Hillsdale, NJ.

Giorgio MAGRI (2020), A principled derivation of Harmonic Grammar, in Allyson ETTINGER, Gaja JAROSZ, and Max NELSON, editors, *Proceedings of the third meeting of the Society for Computation in Linguistics*, Association for Computational Linguistics.

John J. MCCARTHY (2005), Optimal paradigms, in Laura J. DOWNING, T. Alan HALL, and Renate RAFFELSIEFEN, editors, *Paradigms in phonological theory*, Oxford University Press, Oxford.

John J. MCCARTHY and Alan PRINCE (1995), Faithfulness and reduplicative identity, in Jill BECKMAN, Suzanne URBANCZYK, and Laura WALSH DICKEY, editors, *University of Massachusetts occasional papers in linguistics 18: papers in Optimality Theory*, pp. 249–384, GLSA, Amherst.

John J. OHALA (1983), The origin of sound patterns in vocal tract constraints, in Peter F. MACNEILAGE, editor, *The production of speech*, pp. 189–216, Springer-Verlag, New York.

Joe PATER (2009), Weighted constraints in generative linguistics, *Cognitive Science*, 33:999–1035.

Christopher POTTS, Joe PATER, Karen JESNEY, Rajesh BHATT, and Michael BECKER (2010), Harmonic Grammar with linear programming: from linear systems to linguistic typology, *Phonology*, 27(1):1–41.

Alan PRINCE (2002), Entailed ranking arguments, `http://roa.rutgers.edu/files/500-0202/500-0202-PRINCE-0-1.PDF`, manuscript (Rutgers University). Available from the Rutgers Optimality Archive as ROA 500.

Alan PRINCE (2015), One tableau suffices, `http://roa.rutgers.edu/content/article/files/1453_alan_prince_4.pdf`, manuscript (Rutgers University). Available from the Rutgers Optimality Archive as ROA 1250.

Alan PRINCE and Paul SMOLENSKY (1993/2004), *Optimality Theory: Constraint Interaction in generative grammar*, Blackwell, Oxford, `http://roa.rutgers.edu`, original version, Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, April 1993. Available from the Rutgers Optimality Archive as ROA 537.

Fred S. ROBERTS and Barry TESMAN (2005), *Applied combinatorics*, Taylor and Francis Group.

Paul Smolensky and Géraldine Legendre (2006), *The Harmonic Mind*, MIT Press, Cambridge, MA.

Juliet Stanton (2017), *Constraints on the distribution of nasal-stop sequences: an argument for contrast*, Ph.D. thesis, MIT.

Gilbert Strang (2006), *Linear Algebra and its applications*, Thomson Brooks/Cole.

Michel Talagrand (2014), *Upper and lower bounds for stochastic processes*, Springer-Verlag, Berlin, Heidelberg.

Bernard Tranel (1987), *The sounds of French: An introduction*, Cambridge University Press.

Giorgio Magri
0000-0003-0350-9235
magrigrg@gmail.com

SFL (CNRS, University of Paris 8, UPL)

Benjamin Storme
0000-0002-2560-2265
benjastorme@hotmail.com

University of Lausanne

Giorgio Magri and Benjamin Storme (2020), *Constraint summation in phonological theory*, Journal of Language Modelling, 8(2):251–294
https://dx.doi.org/10.15398/jlm.v8i2.216