Phonology **35** (2018) 247–286. © Cambridge University Press 2018
doi:10.1017/S0952675718000040

Positional licensing, asymmetric trade-offs and gradient constraints in Harmonic Grammar[*]

Aaron Kaplan
University of Utah

In Harmonic Grammar, positional licensing interacts with faithfulness constraints in pathological ways: spreading a feature to a licensing position to satisfy positional licensing can incur many faithfulness violations, and if there are sufficiently many such violations, they gang up to block spreading. This problem is solved if positional licensing is recast as a positive constraint that rewards licensed features in proportion to the number of positions they are associated with, thereby countering faithfulness's multiple violations. This proposal provides support for positive constraints, calls into question arguments against gradient constraints and lays the groundwork for a sound theory of positional licensing in Harmonic Grammar.

---

1 Introduction

Optimality Theory (OT; e.g. Prince & Smolensky 1993) and Harmonic Grammar (HG; e.g. Legendre *et al.* 1990) differ in how they adjudicate conflict between constraints. In OT, strict domination ensures that the decision always goes to the higher-ranked constraint, but in HG, as Pater (2009) shows, multiple violations of one or more lower-weighted constraints can gang up on a violation of a higher-weighted constraint, so that the outcome is dictated by the lower-weighted constraint(s). Much research since Pater (2009) has revealed advantages to HG's weighted-constraint approach; I will mention here just one example, because it concerns the constraint families at issue in this paper. Jesney (2011) shows that HG affords an analysis of what she calls 'licensing in

---

[*] E-mail: A.KAPLAN@UTAH.EDU.

I am grateful to the following people for their insightful comments during the development of this work: Abby Kaplan, Wendell Kimper and Rachel Walker, and audiences at AMP 2015, the 37th Annual Conference of the German Linguistic Society (DGfS) and the University of Utah. Thanks also to an associate editor and three anonymous reviewers at *Phonology*, whose comments greatly improved this paper.

multiple contexts' that is unavailable to OT. Positional licensing (PL) constraints (Itô 1986, Goldsmith 1989, Lombardi 1994, Steriade 1995, Zoll 1997, 1998a, b, Crosswhite 2001, Walker 2004, 2005, 2011) penalise elements that do not appear in a designated licensing position, but occasionally languages permit elements restricted in this way to surface in either of two licensing positions. For example, English permits [h] in word-initial and stressed-syllable onsets (see Jesney 2011 for references). Jesney shows that such a pattern arises in HG when two PL constraints (one for each licensing position) are each outweighed by faithfulness, but gang up on faithfulness when they are both violated. In OT the analysis is impossible: with faithfulness undominated, the licensing constraints cannot influence the output even when both are violated. They must outrank faithfulness, but in that case [h] is permitted only where both constraints are satisfied – stressed word-initial syllables – not where it satisfies just one of them.

As this example shows, shifting from OT to HG or *vice versa* can change how particular constraints interact with each other, and Pater (2009) observes that, as a consequence, the two frameworks may require different constraint inventories: constraints that satisfactorily model a phenomenon in one theory may be inadequate in the other, and this may reveal not that the latter framework itself is inadequate, but simply that we're using the wrong constraints. That is in fact a conclusion we can draw from Jesney's study: PL is sufficient for licensing in multiple contexts in HG, but OT requires other constraints (namely positional faithfulness, as Jesney shows).

This paper examines another situation in which HG and OT diverge in the requirements they place on Con, this time arising from a pathological interaction between PL and faithfulness in HG. In addition to confining restricted elements to licensing positions, as with English [h], PL can trigger long-distance assimilation between the licensing position and the position that hosts the restricted element underlyingly. This harmony grants the restricted element membership in the licensor to satisfy PL. For example, in the metaphony system of the Romance variety spoken in Central Veneto, post-tonic [+high] spreads to the stressed syllable, as in (1) (Walker 2005, 2010, 2011).

(1) kalˈseto	masc sg	kalˈsiti	masc pl	'sock'
kanˈtese	1pl	kanˈtisimo	1pl impf subj	'sing'
ˈmovo	1sg	ˈmuvi	2sg	'move'
kanˈtor	masc sg	kanˈturi	masc pl	'choir singer'
ˈordeno	1sg	ˈurdini	2sg	'order'

Under Walker's (2011) OT analysis, the PL constraint License ($[+\text{high}]_{\text{post-tonic}}$, $\acute{\sigma}$), which requires a post-tonic [+high] to coincide with the stressed syllable, drives this assimilation. (Coincidence is defined formally by Zoll 1998a, but for our purposes it is enough to say that α and β coincide if one is autosegmentally associated with, is a feature of, is a constituent of, etc., the other.) Outranking Ident[high], this constraint achieves the desired outcome, with [+high] spreading across any distance to find the licensor. As the distance between trigger and target grows, Ident violations accumulate, but since Ident is dominated by License, this is inconsequential. The stress pattern of Central Veneto seems to preclude more than one vowel appearing between the post-tonic trigger and the stressed syllable (e.g. [ˈurdini]), but we can see the unbounded nature of the system with the schematic form in (2). (I assume other constraints block alternative ways of satisfying License, such as deletion of the final vowel or assimilation that involves skipping the intervening vowels. Some of these alternatives are considered below; see also Walker 2011.)

(2)

/ˈeeee-i/	License([+high]$_{\text{post-tonic}}$,$\acute{\sigma}$)	Ident[high]
a. ˈeeee-i	*!	
☞ b. ˈiiii-i		****

But in HG, distance matters. Since all constraints contribute to a candidate's score, the trigger and target might be so far apart that spreading becomes prohibitively expensive, because of the number of Ident violations it incurs. Such a situation is shown in (3). Here, the weight of License is sufficient to overcome Ident's penalty when two vowels assimilate, but not when three do. The competition between License and Ident involves an asymmetric trade-off, the situation Pater (2009) identifies as giving rise to gang effects in HG: failure to spread violates License just once, but spreading potentially violates Ident many times.

(3) a.

/ˈee-i/	License 5	Ident 2	$\mathcal{H}$
i. ˈee-i	−1		−5
☞ ii. ˈii-i		−2	−4

b.

/ˈeee-i/	License 5	Ident 2	$\mathcal{H}$
☞ i. ˈeee-i	−1		−5
ii. ˈiii-i		−3	−6

(3) exemplifies a language in which metaphony occurs at short distances, but not at long ones. More generally, as we will see, these constraints yield grammars in which, for any arbitrary *n*, harmony applies across *n* intervening positions, but not *n*+1. Languages like this do not seem to exist. For example, Walker's extensive survey of licensing-driven phenomena reports nothing resembling the pattern in (3). In this paper I present revisions to Walker's PL formalism that exclude this pathology. I will argue for two crucial theoretical positions. First, PL must be sensitive to the distance between the trigger and target, assigning not just one violation for an unlicensed feature, but additional violations for positions between the trigger and target that do not themselves assimilate, thus allowing PL to keep pace with the escalating violations of Ident. That is, PL must be gradient in the same way that alignment (McCarthy & Prince 1993), for example, is gradient. The other change draws on Kimper's

(2011) HG-based theory of vowel harmony. Distance-sensitive PL constraints, like the harmony constraints that Kimper examines, introduce their own liabilities. Kimper's solution is the introduction of positive constraints that reward harmony instead of penalising its absence; the same apparatus works for PL. (So in the end, distance-sensitive PL *rewards* each position that harmonises, rather than *penalising* positions that don't.) A third theoretical construct, serialism, will be mentioned occasionally as a remedy for certain other issues that arise, but its application to the distance-sensitive theory of PL just described will not be pursued fully here.

The investigation presented below, then, serves as a case study in the ways in which OT and HG can yield divergent constraint interactions, and it has a number of significant consequences. First, it provides further support for positive constraints. It also reveals important differences between HG and OT in terms of the demands those frameworks place on the proper formulation of PL, and we will end up with a formalism that looks quite different from OT-based PL. Achieving the correct balance between PL and faithfulness requires converting PL into a gradient constraint, a constraint type that has been shown to make erroneous typological predictions (e.g. McCarthy 2003). The success of gradient PL therefore suggests that gradient constraints as a whole are not unambiguously hazardous; the situation is more nuanced, with gradient constraints occasionally having advantages over categorical ones. Finally, this paper represents a first step toward a satisfactory theory of licensing-based phenomena in HG. As will become apparent, especially in §5, the proposal developed here addresses the immediate issues at hand, but constructing a fully articulated theory of PL must be left for future work.

By exploring pathologies like the one illustrated in (3) and a variety of theoretical constructs – categorical *vs.* gradient constraints, negative *vs.* positive constraints and parallel *vs.* serial evaluation – this paper identifies certain essential characteristics of PL and the larger theoretical framework in which PL must be situated. Table I will help us keep track of the consequences of different combinations of these constructs. The top left cell is filled in on the basis of (3), the bottom right cell anticipates the results of the paper and the remainder will be filled in as we go along.

	categorical		gradient	
	negative	positive	negative	positive
parallel	pathologies			
serial				correct prediction

Table I

Consequences of the combinations of various theoretical primitives (version 1).

The paper is structured as follows: § 2 introduces the linguistic patterns at issue, and § 3 presents the pathologies that arise in HG accounts of them. § 4 develops the positive gradient version of PL that avoids these pathologies. § 5 discusses remaining issues, § 6 considers alternatives and § 7 provides discussion and concluding remarks.

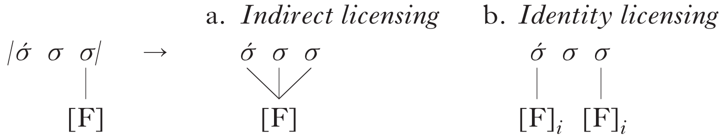
2 Licensing-based patterns

In the literature on licensing-based phenomena, PL constraints come in a variety of forms, each of which, in one way or another, penalises elements that do not appear in a designated licensing position. I take as the basis for the present investigation the theory of PL developed by Walker (2011), which is fairly typical of PL frameworks, and perhaps the most well fleshed-out. Her formalism, presented in a modified form that abstracts away from irrelevant details, is given in (4), where λ is a variable over elements that may be subject to PL (e.g. [+high] in Central Veneto) and π is a variable over possible licensing positions, such as stressed syllables (see Walker 2011 and Kaplan 2015 for studies of both categories).

(4) LICENSE(λ, π)
Assign one violation for each λ that does not coincide with some π.

This constraint family motivates a variety of processes. Here we will examine a subset of what are called OVERWRITE systems in Kaplan (2015): patterns in which PL compels the licensor, such as a stressed syllable, to assimilate to λ. The particular overwrite systems at issue here are ones in which λ spreads to or is copied into π, as in (5a) and (b) respectively. (For a third overwrite pattern, see § 3.1 and § 5.) The term 'overwrite' adopts the perspective of π, whose underlying features are overwritten by these operations. The labels for the licensing configurations in (5) follow Walker (2011).

(5)



Indirect licensing is exemplified by the Central Veneto metaphony in (1): post-tonic high vowels cause both the stressed vowel and intervening vowels to raise. Certain vowels block metaphony, a fact I set aside until § 4.4.1.

Identity licensing appears in Eastern Andalusian, where a word-final [−ATR] vowel triggers harmony in the stressed syllable (Jiménez & Lloret 2007, Lloret & Jiménez 2009). The [−ATR] vowels are transcribed in a number of ways in the literature; I adopt the symbols [ɪ ɛ ʊ ɔ] for the [−ATR] counterparts of [i e u o] respectively, and [æ̞ a̞] for the fronted and non-fronted lax counterparts of [a]. Word-final /s/ is debuccalised or deleted in Eastern Andalusian – a process often called /s/-aspiration – and Jiménez & Lloret (2007) and Lloret & Jiménez (2009) argue that /s/'s [spread glottis] feature is preserved on the preceding vowel in the form of [−ATR]. Aspiration, word-final laxing and ATR harmony are all visible in (6). Where available, morphologically related forms lacking the context for these processes are provided for comparison. The properties of these word-final vowels and the harmony they trigger is the subject of considerable disagreement in the literature; I follow Sanders (1998), Jiménez & Lloret (2007) and Lloret & Jiménez (2009) (see Sanders 1998 for a discussion of other studies). I will not present an explicit account of /s/-aspiration or word-final laxing here, nor will I provide an account of the fronting of /a/ visible in examples like *asas*.

(6)	*nenes*	ˈnɛnɛ	‘babies’	*nene*	ˈnene	‘baby’
	monos	ˈmɔnɔ	‘monkeys’	*mono*	ˈmono	‘monkey’
	asas	ˈa̞sæ̞	‘handles’	*asa*	ˈasa	‘handle’
	lejos	ˈlɛhɔ	‘far’			
	mes	ˈmɛ	‘month’			
	tos	tɔ	‘cough’			
	mis	mɪ	‘my (PL)’	*mi*	ˈmi	‘my (SG)’
	tus	tʊ	‘your (PL)’			
	tesis	ˈtɛsɪ	‘thesis’			
	pesos	ˈpɛsɔ	‘weights’	*peso*	ˈpeso	‘weight’
	bocas	ˈbɔkæ̞	‘mouths’	*boca*	ˈboka	‘mouth’
	tienes	ˈtjɛnɛ	‘you have’	*tiene*	ˈtjene	‘he/she has’

If the final and stressed vowels are not syllable-adjacent, the intervening vowels optionally harmonise, as in (7). If there is more than one such vowel, they harmonise (or not) uniformly.

(7)	*treboles*	ˈtɾɛβolɛ ~ ˈtɾɛβɔlɛ	‘clovers’
	cómetelos	ˈkɔmetelɔ ~ ˈkɔmɛtɛlɔ *ˈkomɛtelɔ, *ˈkometɛlɔ	‘eat them (for you)!’

The variants without harmony on the intervening vowels illustrate identity licensing. Two other properties of Eastern Andalusian harmony are notable: it optionally extends to pre-tonic vowels, in which case all vowels – pre-tonic and post-tonic – must harmonise, as in (8a), and while high vowels undergo laxing word-finally, they resist harmony when stressed, as in (b). I will not address pre-tonic harmony in detail here; in Kaplan (2017), which is summarised in §4.4.2, I show that the formalism developed below extends to this harmony. I also set the harmony-resistant high vowels aside until §4.4.2. Throughout, the treatment of Eastern Andalusian closely follows Jiménez & Lloret (2007), Lloret & Jiménez (2009) and Walker (2011).

(8) a.	*momentos*	moˈmɛntɔ ~ mɔˈmɛntɔ	‘instants’
	reloj	reˈlɔ ~ rɛˈlɔ	‘watch’
	relojes	reˈlɔhɛ ~ rɛˈlɔhɛ	‘watches’
	monederos	moneˈðɛɾɔ ~ mɔnɛˈðɛɾɔ *mɔneˈðɛɾɔ, *monɛˈðɛɾɔ	‘purses’
	cojines	koˈhinɛ ~ kɔˈhinɛ	‘pillows’
	cotillones	kotiˈʒɔnɛ ~ kɔtiˈʒɔnɛ	‘cotillions’
	recógelos	reˈkɔhelɔ ~ reˈkɔhɛlɔ ~ rɛˈkɔhɛlɔ *rɛˈkɔhelɔ	‘pick them’

b.	*crisis*	ˈkɾisɪ	‘crisis’
	muchos	ˈmuʃɔ	‘many’
	mios	ˈmiɔ	‘mine (PL)’

Harmony in Central Veneto and Eastern Andalusian is driven by PL. In OT, PL dominates faithfulness (IDENT[high] for Central Veneto, IDENT[ATR] for Eastern Andalusian). Identity licensing satisfies PL at minimal cost to faithfulness, because only the licensor assimilates. The resulting discontiguous harmony domain violates *DUPLICATE in Walker’s (2011) system. Indirect licensing yields a contiguous harmony domain at the expense of faithfulness. Thus the two patterns are distinguished by the ranking of *DUPLICATE and faithfulness, as (9) demonstrates (the tableaux are minimally modified from Walker 2011).

(9) a.

/ˈordeni/	LICENSE([+high]$_{\text{post-tonic}}$,σ́)	*DUP	IDENT[high]
i. ˈordeni	*!		
☞ ii. ˈurdini			**
iii. ˈurdeni		*!	*

b.

/ˈtɾeβol-es/	LICENSE([−ATR],σ́)	IDENT[ATR]	*DUP
i. ˈtɾeβolɛ	*!	*	
☞ ii. ˈtɾɛβolɛ		**	*
iii. ˈtɾɛβɔlɛ		***!	

As before, I set aside other ways of satisfying PL (but see §5 for discussion). Because of strict domination, each ranking in (9) always favours the language-appropriate licensing configuration. As illustrated in (3), and shown more fully in the following section, this is not true in HG. The aim of this paper is to probe and correct the resulting pathologies.

3 Positional licensing and its pathologies

Recasting (4) for HG yields (10).

(10) License(λ, π)
Assign −1 for each λ that does not coincide with some π.

The ranking in (9a) renders Ident powerless to stop assimilation in both the licensor and any intervening syllables, but in HG the accumulation of Ident violations eventually blocks harmony, regardless of constraints' weights (unless $w(\text{License}) = \infty$; see §6.1). For example, with the weights in (11), harmony is blocked when it must cross two intervening vowels. Here and in many subsequent tableaux, I use schematic, idealised forms that allow us to better visualise the predicted patterns, and avoid for now language-particular complications. These idealisations are based on a common type of metaphony, where a final high vowel triggers raising of the stressed vowel and possibly also the intervening vowels. All possible winners are indicated; when there is more than one, the conditions yielding each are indicated after the pointing finger.

In (11), identity licensing – candidate (c) – violates *Duplicate and Ident once each, giving a score of −6 (e$_0$ and i$_0$ indicate any number of unharmonised and harmonised vowels respectively, including zero).

(11)

/ˈee$_0$-i/	License 5	*Dup 4	Ident 2	$\mathcal{H}$
☞ (n>2.5) a. ˈee$_0$-i	−1			−5
☞ (n<2.5) b. ˈii$_0$-i			−n	−2n
c. ˈie$_0$-i		−1	−1	−6

Under these weights, this candidate is always worse than the faithful form, whose lone License violation gives a score of −5. The score for indirect licensing is proportional to the number of harmony targets, where n is the total number of positions that assimilate in indirect licensing (the licensor plus any intervening positions). The score for indirect licensing, then, is $-w_I \times n$, where w_I is the weight of Ident. (As we revise PL, the way in which harmony scores are projected will change, but n will remain central to this calculation.) When $-w_I \times n$ is less than the faithful form's score (here, when $n > 2.5$), indirect licensing becomes too costly, and the faithful form wins. I will call this the No Distant Licensing pathology, because harmony occurs only when there is a short distance between the trigger and the licensor.

The source of the pathology is the asymmetric trade-off between License and Ident. Pater (2009: 1017) calls this kind of asymmetric trade-off 'unbounded': 'satisfaction of one constraint can require a potentially unbounded number of violations of another'. See O'Hara (2016) for a similar unbounded trade-off in harmony not driven by PL and Legendre *et al.* (2006) for an example from metrical phonology. Harmony occurs iff the inequality in (12) holds.

(12) $n \times w(\textsc{Ident}) < w(\textsc{License})$

From this, we can deduce that the greatest distance across which indirect licensing may occur – the maximum number of positions n that may assimilate – is as in (13).[1]

(13) $n < w(\textsc{License}) / w(\textsc{Ident})$

No matter the constraint weights, there is always some n for which IDENT violations overwhelm the lone violation of LICENSE. Consequently, the No Distant Licensing pathology is not an obscure pattern that arises only under certain circumstances; the system predicts that indirect licensing always works this way. But languages like this appear to be unattested: I know of no language that shows licensing-driven harmony across an arbitrarily short distance but not across longer distances.

The No Distant Licensing pathology contrasts with phenomena that show distance-based decay (Hayes & Londe 2006, Zymet 2015), wherein the frequency with which segments interact decreases as they become further apart. For example, Zymet argues that rounding dissimilation in Malagasy (Parker 1883) occurs less frequently as more syllables appear between round vowels. The passive imperative suffix /-u/ surfaces as [i] when attached to a stem that contains /u/: [tuv-i] ‘fulfil’. Zymet finds that this dissimilation is nearly exceptionless (it occurs at rate of 0.99) when the two vowels are in adjacent syllables. Exceptions become more frequent as more syllables intervene: with one intervening syllable, the proportion of dissimilatory forms is 0.51, with two it is 0.13, etc.

The No Distant Licensing pathology predicts not a gradual decay with distance, but rather a sharp boundary between short distances, with uniformly regular harmony, and long distances, with no harmony. It is also distinct from sibilant harmony in Navajo (Martin 2005), where adjacency is crucial: sibilants in adjacent syllables show harmony more frequently than more distant sibilants. Under the No Distant Licensing pathology, the boundary can occur at any distance – adjacency has no special privilege. Furthermore, Navajo exhibits a contrast between more and less frequent harmony, whereas the No Distant Licensing pathology predicts a complete cessation of harmony at longer distances.

A related pathology arises under different weights. In (14), indirect licensing gives way to identity licensing at longer distances. This pattern, which I will call the IDENTITY AT A DISTANCE pathology, emerges when LICENSE outweighs the combined weights of IDENT and *DUPLICATE so that it can compel a single violation of both of those constraints. Whereas previously identity licensing was always less harmonic than no harmony, here the opposite holds, so that when indirect licensing

---

[1] If the ratio of the weights of LICENSE and IDENT is a whole number x, then when $n = x$, the indirect licensing candidate and the faithful candidate tie on the relevant constraints. I set this situation aside.

becomes too costly (when $n > 3$), identity licensing emerges. Systems like this are unattested.

(14)

/ˈee$_0$-i/	License 7	*Dup 4	Ident 2	$\mathcal{H}$
a. ˈee$_0$-i	−1			−7
☞ (n<3) b. ˈii$_0$-i			$-n$	$-2n$
☞ (n>3) c. ˈie$_0$-i		−1	−1	−6

Figure 1 presents the problem. Each line shows the harmony score for one of the candidate types represented in (11) and (14) as a function of *n*. In the two graphs, the scores for the faithful form and identity licensing are constant, but the score for indirect licensing is inversely proportional to *n*. The intersection of the function for indirect licensing and one of the other functions – where the lines meet in the graphs – marks the point at which indirect licensing ceases to be most harmonic. By changing the weights we can manipulate the horizontal lines' *y*-intercepts and the slope of the line for indirect licensing, but with non-zero weights the functions always intersect. If they do so when $n > 1$ – a necessity under weights that favour indirect licensing for any configuration at all, assuming only positive weights – indirect licensing occurs at short distances only.[2] Whether indirect licensing gives way to the faithful form or identity licensing essentially depends on *Duplicate: is its weight sufficient to render identity licensing worse than no harmony? Eliminating the pathologies, then, means devising constraints whose harmony functions in the final column in each of the tableaux above do not intersect in the domain $n > 1$ for any weights.[3]

The remainder of this section provides further support for the claim that the patterns discussed above warrant attention. §3.1 reinforces the argument that these patterns are indeed pathological, and §3.2 shows that they arise not just under parallel HG, but also under serial HG.

3.1 Attested and unattested patterns

The pathologies described above involve arbitrary limits on the reach of positional licensing. While there is no reason to believe that languages respect some maximum-distance limitation, the available data actually

---

[2] Recall that *n* is the number of positions that harmonise in indirect licensing, so only positive integers are meaningful. And if $n = 1$, there are no intervening positions, and the distinction between indirect and identity licensing is uninteresting (though see Walker 2011 for reasons to think that in such situations languages favour indirect over identity licensing).

[3] This condition is reminiscent of Prince's (2003) Anything Goes systems. An Anything Goes system is one in which an OT result can be replicated in HG with any weights that mirror the OT ranking (i.e. if $C_i \gg C_j$ then $w(C_i) > w(C_j)$); non-Anything Goes systems impose more specific weighting requirements. Here we need constraints that interact so that for any weights they produce a single OT result, regardless of *n*.

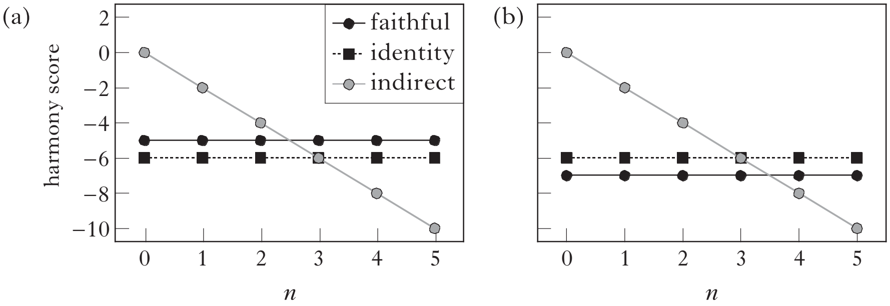


Figure 1
Harmony scores for the candidates and weights in (a) (11) and (b) (14).

exemplifies only surprisingly short distances. This is likely because metaphony systems are most abundant in Romance varieties, whose prosodic systems locate stress (the target) near the right edge of the word (the trigger). For example, I know of no data from Central Veneto with more than one intervening vowel. To my knowledge, the largest distance across which metaphony-like harmony is attested is the two intervening vowels we saw in Eastern Andalusian. However, there are at least two other examples of licensing-based systems that exhibit greater distances between the trigger and target. In Esimbi (Stallcup 1980a, b, Hyman 1988), non-high vowel features move to the initial syllable, so that, as illustrated in (15), roots surface with high vowels, while prefixes show a range of heights. For example, the stem [si] underlyingly contains (the features for) a mid vowel, but those features surface on the prefix. Whereas PL was satisfied in Central Veneto and Eastern Andalusian via harmony, here this is accomplished by transferring the relevant features to the licensor. Walker (2011) calls this DIRECT LICENSING: the restricted element surfaces exclusively in the licensor. Of relevance here is that the data show up to three positions that host unlicensed features underlyingly.

(15) *underlying stem* V	*infinitive*		*class 9 sg*	
/i/	u-bini	‘dance’	ì-ʤìmì	‘back’
/u/	u-suhuru	‘crouch’	ì-sù	‘fish’
/e/	o-si	‘laugh’	è-gbì	‘bushfowl’
/o/	o-zumu	‘dry up’	è-nùnù	‘bird’
/ə/	o-tɨnɨ	‘refuse’	è-kpɨ̀sɨ̀	‘rock’
/ɛ/	ɔ-rini	‘be poor’	ɛ̀-njìmì	‘animal’
/ɔ/	ɔ-zumulu	‘wither’	ɛ̀-fumù	‘hippopotamus’
/a/	ɔ-sɨmbɨrɨ	‘scatter’	ɛ̀-kɨ̀rɨ̀	‘headpad’

Classical Mongolian’s rounding harmony involves domains of similar lengths (Poppe 1954, 1955, Walker 2001). Here, ‘non-high round vowels are permitted in non-initial syllables of the root only when all preceding syllables contain non-high round vowels' (Walker 2011: 99). The language's vowel inventory is [i y e ø a o u]; the possible combinations of round and unrounded vowels is illustrated in (16a). (Backness harmony is also visible.) In contrast, the illicit forms in (b) do not obey the restriction on rounding.

(16) a.	nøkør	'friend'	b.	*nekør
	ølø	'grey'		*begørø
	moŋɣol	'Mongol'		*namoɣodqo
	qomoɣol	'horse-dung'		*nomaɣodqa
	møren	'river'		
	kømøske	'eyebrow(s)'		
	bøgere	'kidney'		
	qola	'far, distant'		
	olan	'many'		
	nomoɣodqa	'to tame'		

In Kaplan's (2015) terminology, this is a PRESERVATION system: the licensor remains faithful, and other positions either harmonise with it (creating configurations like (5a) or (5b)) or lose their unlicensed features altogether. In Classical Mongolian, if [+round] on a non-high non-initial vowel is not shared by an initial non-high vowel, that feature is eradicated; there is no active assimilation in the language in either direction (Walker 2001). The words in (16a) show up to three non-initial vowels, so like Esimbi we have evidence for up to three positions that may host unlicensed features. In neither language is word length a factor in the resolution of PL violations, yet distance-based pathologies similar to the ones presented above also arise here. Escaping one PL violation in these languages involves removing features from non-licensing positions, which entails potentially many Ident violations, setting up the same asymmetric trade-off we saw above – e.g. in (11) – the difference here being that faithfulness violations arise from moving or eliminating the restricted feature rather than harmonising it. In no language that I am aware of does the number of faithfulness violations matter in these kinds of systems.

These examples give us sufficient reason to be confident that licensing systems that block harmony at (relatively) long distances do not exist: traditional PL incorrectly predicts analogues of Eastern Andalusian, Esimbi and Classical Mongolian that prohibit or alter harmony at the greatest distances provided by those languages.

Two other considerations are worth bearing in mind. First, the constraint interactions described above predict that *all* licensing-based systems have a maximum distance, a claim for which there is clearly no basis. (It would seem suspiciously coincidental that constraint weights and word length conspire to make the maximum distance invisible in all languages.) Second, the Identity at a Distance pathology presents not just a maximum distance, but a distance-based switch from one licensing pattern to another, a system for which there is no evidence within the typology of licensing-based phenomena.

For these reasons I conclude that the systems claimed to be pathological in this section are indeed impossible. The asymmetry between LICENSE and IDENT is the source of these pathologies, so to eliminate them we must eliminate this asymmetry. (Adding another constraint to counterbalance IDENT is inadequate; see §6.) Consequently, two broad strategies are available: we can reformulate PL so that it assigns violations in proportion to the distance between the trigger and target, or we can change IDENT so that it does not. I will pursue the first option. In particular, I will argue for the formulation of PL in (17), though we will consider other interim formulations along the way. §6.3 shows that changing IDENT is not viable.

(17) LICENSE(λ, π)

Assign +1 for each λ that coincides with some π. For each λ that coincides with some π, assign +1 for each additional position that λ coincides with.

This formalism differs from traditional PL in two significant ways. It rewards licensed features instead of penalising unlicensed features, and that reward is proportional to the number of positions that assimilate. In §4 I argue that both properties are essential, but first I examine the behaviour of PL under serialism. There are two reasons for this excursus on serialism. First, serialism appears to offer a solution to these pathologies, because, with only one change allowed at a time, it prevents faithfulness from ganging up on PL. We will see, though, that the No Distant Licensing pathology arises nonetheless – the pathology occurs in both parallel and serial evaluation, and is therefore a pervasive problem. And second, although serialism does not escape this pathology, it has a crucial role to play, for reasons that will become clear in §4. Consequently, it is important to clearly delineate what serialism does and does not do for the issues at hand.

3.2 The pathologies in Serial Harmonic Grammar

The ostensible advantage of serialism is that it requires competitions between constraints to play out locally (McCarthy 2006, Pater *et al.* 2007): since only one change can occur on any step, at any moment the lone violation of PL must contend with a single faithfulness violation, not the army of violations we considered above. It might therefore appear that it avoids the distance-based pathologies. However, serialism merely replaces one pathological constraint interaction with another.

Kimper (2012) shows how the long-distance effects of PL can be achieved serially; see also Walker (2010) and Kaplan (2011) for more on PL's behaviour under serialism. The path to indirect licensing begins with harmony on the licensor. This step is illustrated in (18), using the categorical version of PL given in (10).

(18) *Step 1*

/ˈeee-i/	License 6	Ident 1	$\mathcal{H}$
a. ˈeee-i	−1		−6
☞ b. ˈiee-i		−1	−1
c. ˈeei-i	−1	−1	−7

This harmony satisfies PL (and it is necessarily the first step, because harmony elsewhere serves no purpose, as candidate (c) shows), and obviously some other constraint must trigger harmony on the non-licensing positions. Kimper shows that *Duplicate is ineffective here: the winner in (18) violates *Duplicate once, as do the candidates representing the next step toward complete harmony, [ˈiie-i] and [ˈiei-i]. All intervening vowels must harmonise before the *Duplicate violation is eliminated, and therefore the pay-off comes only after a series of steps, a situation McCarthy (2008) shows cannot be produced in a gradual framework (see also Walker 2010). In (19), ☜ and 👎 mark the intended and actual winners respectively.

(19) *Step 2*

/ˈiee-i/	License 6	*Dup 2	Ident 1	$\mathcal{H}$
👎 a. ˈiee-i		−1		−2
☜ b. ˈiei-i		−1	−1	−3

Instead, Kimper introduces a *Skip(V) family of constraints that assigns one violation for each vowel that intervenes between the two halves of the harmony domain. Now [ˈiie-i] and [ˈiei-i] are more harmonic than [ˈiee-i], as shown in (20).[4]

(20) *Step 2*

/ˈiee-i/	License 6	*Skip(V) 2	Ident 1	$\mathcal{H}$
a. ˈiee-i		−2		−4
☞ b. ˈiie-i		−1	−1	−3
☞ c. ˈiei-i		−1	−1	−3

The resolution of the tie in (20) is immaterial; whichever candidate wins, the remaining vowel harmonises on the next step, as illustrated in (21), assuming the input /ˈiei-i/.

---

[4] In the current situation, *Skip(V) does not change the outcome of (18), where the winning form's score becomes −5, but this is not always the case, as we will see in (22).

(21) *Step 3*

/ˈiei-i/	LICENSE 6	*SKIP(V) 2	IDENT 1	$\mathcal{H}$
a. ˈiei-i		−1		−2
☞ b. ˈiii-i			−1	−1

Convergence occurs at Step 4. However, this system suffers from the No Distant Licensing pathology. When Step 1 – harmony in the licensor – incurs sufficiently large numbers of *SKIP(V) violations, harmony fails. Using the weights from the foregoing derivation, this occurs when there are three intervening vowels, as in (22).

(22) *Step 1*

/ˈeeee-i/	LICENSE 6	*SKIP(V) 2	IDENT 1	$\mathcal{H}$
👎 a. ˈeeee-i	−1			−6
☜ b. ˈieee-i		−3	−1	−7

It is possible that this problem could be alleviated by the proposal in Pater *et al.* (2007) that a markedness constraint assigns a violation only when the operation that produces a candidate from an input alters the markedness constraint's locus of violation (e.g. NOCODA penalises a particular coda just when the operation that gave rise to the candidate affected that coda), but, as those authors acknowledge, their proposal remains far too imprecise to apply to contexts like the current one. (What is *SKIP(V)'s locus of violation in (22): one or both of the harmonised vowels?; one or more of the intervening vowels?; and has that locus been changed in the unfaithful candidate?)

On its own, then, serial HG does not provide a solution to distance-based pathologies. It replaces the asymmetry between PL and faithfulness with one between PL and *SKIP(V): alleviating one violation of the former can trigger many violations of the latter. This does not mean that serialism and PL are incompatible, however. We will see in §4.2 that serialism is in fact advantageous in its own way; the point here is simply that both parallel and serial evaluation suffer from the distance-based pathologies under traditional PL. We can therefore update our chart to give Table II.

4 Positional licensing for HG

This section develops a PL formalism that escapes the distance-based pathologies. I argue that PL must be sensitive to the distance between the trigger and target (§4.1), and that it must be a positive constraint (§4.2).

	categorical		gradient	
	negative	positive	negative	positive
parallel	pathologies			
serial	pathologies			correct prediction

Table II

Consequences of the combinations of various theoretical primitives (version 2).

4.1 Distance-sensitive licensing

As we have seen, the penalty assessed by faithfulness for indirect licensing escalates with the number of intervening positions. If PL's incentive for harmony is to keep pace, the penalty it assigns to unlicensed features must escalate similarly. The revision of (10) in (23), which represents just a first step, achieves this. For each vowel that is penalised by IDENT when it harmonises, LICENSE now assigns a penalty when that position does not harmonise.[5]

(23) LICENSE(λ, π) (revised; version 1)

Assign −1 for each λ that does not coincide with some π, and −1 for each syllable that intervenes between λ and the nearest π.

The intuition here is that PL often doesn't merely trigger an interaction between a feature's underlying host and the licensor – it can also drag positions between those elements into the interaction. Either we capture this with a constraint like *DUPLICATE, as Walker (2011) does, or we build it into PL itself. The argument here is that the latter provides a solution for the pathologies at issue.

As (24) shows, PL now assigns as many violations for failure to assimilate as IDENT does for indirect licensing, and the asymmetry that gave rise to the No Distant Licensing pathology is gone. When w(IDENT) < w(LICENSE), as in (24), the faithful form is always worse than indirect licensing. If IDENT outweighs LICENSE, the opposite is true. The contest between faithfulness and indirect licensing now depends entirely on constraint weights, not the distance between the trigger and target.

(24)

/ˈee$_0$-i/	LICENSE 3	IDENT 2	$\mathcal{H}$
a. ˈee$_0$-i	$-n$		$-3n$
☞ b. ˈii$_0$-i		$-n$	$-2n$

---

[5] Measuring distance by syllables seems adequate for the vocalic licensing systems I am aware of, but I leave open the possibility that this is not the right metric, or that the metric may vary across languages.

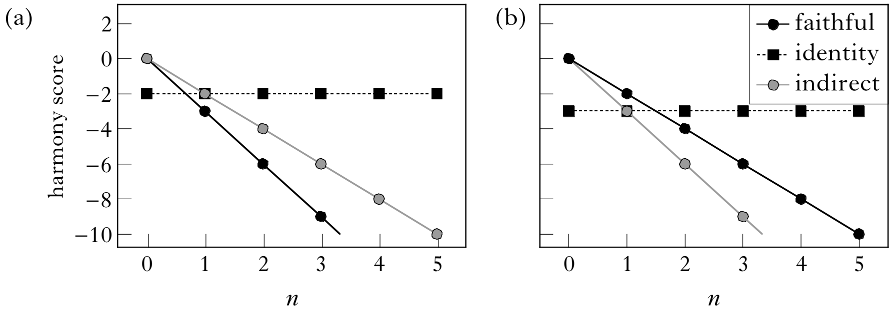


Figure 2

Harmony scores for the candidates and weights in (a) (24) and (b) (25).

The functions for the faithful form and indirect licensing now intersect at $n = 0$ (see Fig. 2), and therefore the superiority of one candidate over the other is constant in the domain of interest. However, as Fig. 2 shows, regardless of which of those candidates is favoured, it gives way at long distances to identity licensing. Fig. 2a shows the Identity at a Distance pathology, and Fig. 2b shows a new pattern, in which assimilation occurs only at long distances. This outcome is also illustrated in (25). Here and subsequently I omit *DUPLICATE to make the interaction between PL and faithfulness more transparent. *DUPLICATE only adds an invariant penalty to identity licensing (so in the present example it simply alters the point at which this candidate is most harmonic), and we will see that its role in favouring indirect over identity licensing is ultimately taken over by PL.

(25)

/ˈee$_0$-i/	IDENT 3	LICENSE 2	$\mathcal{H}$
☞ (n<1.5) a. ˈee$_0$-i		$-n$	$-2n$
☞ (n>1.5) b. ˈie$_0$-i	-1		-3
c. ˈii$_0$-i	$-n$		$-3n$

Identity licensing's score is static, and therefore the situations in Fig. 2 arise under any weights. Either faithfulness or indirect licensing wins at short distances, and at some point their accumulated violations become too much; every language is predicted to show identity licensing if words of sufficient length can be found. The situation is remedied if LICENSE's penalty for failure to target the intervening positions persists in identity licensing – i.e. if each unharmonised vowel in, say, [ˈieee-i] incurs a penalty of −1 from LICENSE. Let us provisionally make that change here; I do not offer a formal definition of this version of PL, because when PL is recast as a positive constraint in §4.2, this aspect of the formalism will look quite different.

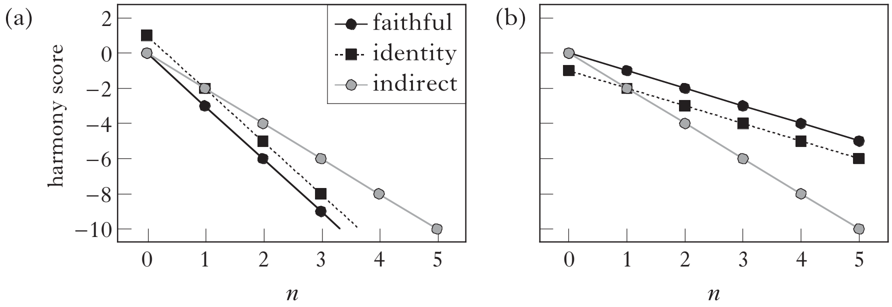


Figure 3
Harmony scores for the candidates and weights in (a) (26a) and (b) (26b).

Figure 3 shows that the pathologies from Fig. 2 cannot arise, a result I verified with OT-Help (Staubs *et al*. 2010). The tableaux in (26) show the weights used to produce these graphs. Figure 3 shows that under the current arrangement, the slopes of the lines for the faithful form and identity licensing are identical – both depend on the weight of LICENSE. Which one comes out ahead depends on whether LICENSE outweighs IDENT. If w(LICENSE) > w(IDENT), identity licensing scores better, while the opposite favours no harmony.

(26) a.

/ˈee$_0$-i/	LICENSE 3	IDENT 2	$\mathcal{H}$
i. ˈee$_0$-i	$-n$		$-3n$
ii. ˈie$_0$-i	$-(n-1)$	-1	$-2-3(n-1)$
☞ iii. ˈii$_0$-i		$-n$	$-2n$

b.

/ˈee$_0$-i/	LICENSE 3	IDENT 4	$\mathcal{H}$
☞ i. ˈee$_0$-i	$-n$		$-3n$
ii. ˈie$_0$-i	$-(n-1)$	-1	$-4-3(n-1)$
iii. ˈii$_0$-i		$-n$	$-4n$

Indirect licensing always intersects the identity-licensing function at $n=1$, regardless of constraint weights. We are not interested in situations where $n \leq 1$ (see note 2), so if indirect licensing beats identity licensing at one distance, it does so at all distances, and *vice versa*. Which of those two options scores better depends on their slopes, which are equal to $-w$(IDENT) and $-w$(LICENSE) respectively: under w(LICENSE) > w(IDENT), indirect licensing has a higher score, and the opposite relationship favours identity licensing.

Unfortunately, the implication of the two preceding paragraphs is that identity licensing is collectively harmonically bounded (Samek-Lodovici

& Prince 1999, 2005). It loses to indirect licensing if w(License) > w(Ident): to take the weights in (26a) as an example, identity licensing's score from License $(-2-3(n-1))$ is worse than $-2n$ for $n>1$. And it loses to the faithful form when w(Ident) > w(License): in (26b), $-4n < -3n$. (*Duplicate, which only penalises identity licensing, does not change this result.) Correcting this state of affairs turns out to be a serendipitous by-product of recasting PL as a positive constraint, so I delay further discussion until that move is made.

Incidentally, the current, gradient version of PL escapes the distance-based pathologies in a parallel evaluation, but not in a serial evaluation. Illustrative tableaux are provided in (27). In (a), the first step of harmony is successful: eliminating one violation of License is worth the cost of one Ident violation and two *Skip(V) violations, because w(License) > 2w(*Skip(V)) + w(Ident). But in the longer form in (b), the same harmony is blocked by the additional *Skip(V) violation. The first step in any derivation removes one violation of License at the cost of potentially many *Skip(V) violations – an asymmetric trade-off.

(27) a. *Step 1*

/ˈeee-i/	License 6	*Skip(V) 2	Ident 1	$\mathcal{H}$
i. ˈeee-i	−3			−18
☞ ii. ˈiee-i	−2	−2	−1	−17

b. *Step 1*

/ˈeeee-i/	License 6	*Skip(V) 2	Ident 1	$\mathcal{H}$
👎 i. ˈeeee-i	−4			−24
☜ ii. ˈieee-i	−3	−3	−1	−25

The harmonic-bounding issue notwithstanding, distance-sensitive PL, in which unassimilated intervening positions are penalised, represents significant progress. With both PL and faithfulness assigning penalties according to the distance that harmony crosses or fails to cross, violations of one cannot gang up on the other, and the No Distant Licensing and Identity at a Distance pathologies do not emerge. We can update our summary as in Table III. On its own, gradience sidesteps the distance-based pathologies (in a parallel framework), but undergenerates by excluding identity licensing.

The next section builds on this progress by addressing remaining defects, most saliently the harmonic bounding of identity licensing. Additionally, distance-sensitive PL is susceptible to problems that plague other gradient constraints, and it counterintuitively penalises unharmonised intervening positions under identity licensing, despite the fact that this configuration achieves the central aim of PL – the coincidence

	categorical		gradient	
	negative	positive	negative	positive
parallel	pathologies		no identity licensing	
serial	pathologies		pathologies	correct prediction

Table III
Consequences of the combinations of various theoretical primitives (version 3).

of an element and its licensor. By reframing PL in positive terms, we address all of these concerns.

4.2 Positive licensing

The previous section recast PL as a gradient constraint, rendering it similar to alignment (McCarthy & Prince 1993), for example. Such constraints suffer from well-known defects (Wilson 2003, McCarthy 2004, 2010, Kimper 2011). For example, they can trigger deletion instead of assimilation of disharmonic segments. Recall that in Central Veneto, post-tonic [+high] spreads through intervening vowels to reach the stressed syllable, as in (1). But with the weights in (28), those vowels delete – but they only do so in metaphony contexts, where License provides the necessary incentive. Such a language is implausible.

(28)

/ˈordeni/	License([+high]$_{\text{post-tonic}}$,σ́) 3	Ident 2	Max 1	$\mathcal{H}$
a. ˈordeni	−2			−6
b. ˈurdini		−2		−4
☞ c. ˈurdni		−1	−1	−3

Second, gradient constraints can block epenthesis that would result in a greater penalty from the gradient constraint. In the current version of PL, the intervening vowels in Eastern Andalusian's identity-licensing variants (e.g. [ˈkɔmetelɔ] in (7)) are penalised for their lack of harmony. Any additional epenthetic vowels would add to this penalty. If that new penalty is worse than the penalty for failure to epenthesise, epenthesis fails. We can see this with Eastern Andalusian′, a hypothetical language with harmony identical to Eastern Andalusian and an additional prohibition on codas. With the constraint weights in (29), the medial coda in /ˈtɾeβtol-es/ is retained, because the penalty assessed by License and Dep for epenthesis (candidate (29a.iii)) is too steep for epenthesis to be viable in a harmony context. In contrast, when harmony is not an issue, epenthesis occurs, as in (29b).

(29) a.

/ˈtɾeβtol-es/	License([−ATR],σ́) 4	Ident 3	NoCoda 2	Dep 1	$\mathcal{H}$
i. ˈtɾeβtolɛ	−2	−1	−1		−13
☞ ii. ˈtɾɛβtolɛ	−1	−2	−1		−12
iii. ˈtɾeβotolɛ	−3	−1		−1	−16
iv. ˈtɾɛβotolɛ	−2	−2		−1	−15
b. /ˈtɾeβtol/					
i. ˈtɾeβtol			−1		−2
☞ ii. ˈtɾeβotol				−1	−1

I am aware of no language that exemplifies the patterns in (28) and (29).

Serial frameworks like Harmonic Serialism (Prince & Smolensky 1993) avoid the deletion-under-metaphony pattern (McCarthy 2010), but not the blocked-epenthesis pattern (Kimper 2011). Instead, Kimper argues for positive harmony constraints as the proper device for eliminating both pathologies. (Positive constraints themselves require serialism, as discussed in Kimper 2011 and later in this section, but strictly in terms of the pathologies currently at issue it is superfluous once positive constraints are adopted.)

Positive constraints reward candidates that have unmarked properties instead of penalising ones that have marked properties. Recasting the current formulation of PL as a positive constraint, we arrive at (30).

(30) License(*λ*, *π*) (revised; version 2)

Assign +1 for each *λ* that coincides with some *π*, and +1 for each additional position that coincides with *λ*.

Instead of penalising unlicensed elements, (30) rewards those that are licensed. It also rewards each additional position such elements are associated with – this preserves the gradient nature of the constraint from §4.1. But there are questions here that did not arise under the negative formulation: under what conditions is this extra reward assigned, and what do we do about the underlying host of *λ*?

To elaborate on the first question, does License(*λ*, *π*) assign rewards for *λ*'s association with non-licensing positions universally, or only when *λ* is licensed? That is, do candidates like [ˈee-i] and [ˈei-i] receive rewards? If so, when the licensor cannot harmonise, PL motivates spreading *λ* as close as possible to it. Licensing-based systems do not seem to work like this. For example, stressed /a/ does not raise in Central Veneto (see §4.4.1): [ˈangol-i] 'angle (masc sg/pl)', *[ˈungul-i]. The intervening vowel also fails to raise in this example, indicating that the reward for *λ*'s appearance in positions besides the licensor must be contingent upon its appearance in the licensor itself. Aside from the empirical evidence just provided, this arrangement better reflects the primary goal of PL, which is to achieve coincidence between some element and a particular position, not general harmony.

To be explicit on this point, we can slightly revise the definition in (30), to give (31), anticipated in (17) above.

(31) License(λ, π) (final version)

Assign +1 for each λ that coincides with some π. For each λ that coincides with some π, assign +1 for each additional position that λ coincides with.

As for the second question, (31) does not distinguish the underlying host of λ from positions that undergo harmony; both positions receive +1. Consequently, we do not have exactly the same symmetry between gradient License and Ident that we had before, where, as illustrated in (26), License and Ident penalised the same set of positions, namely the licensor and any intervening positions. Now, in addition to potentially rewarding that set, License also rewards each position that faithfully retains the harmonising feature from the input. The result is shown schematically in (32) for forms of various lengths, where the underlined vowels are the underlying hosts.

(32)

	License	Ident
a. ˈi-i̲	+2	−1
b. ˈii-i̲	+3	−2
c. ˈiii-i̲	+4	−3

We could maintain the earlier symmetry by excluding the underlying host from License's reward, but doing so is not trivial. Distinguishing the original host from harmonisers would require some notion of domain heads (e.g. Cole & Kisseberth 1994, McCarthy 2004), or enabling PL to 'see' underlying associations, power that Kaplan (2008) argues against. Alternatively, we could simply subtract 1 from the reward License assigns, though this seems blatantly ad hoc and assumes exactly one underlying host. It turns out that keeping things as they are, with underlying hosts being rewarded, alleviates identity licensing's harmonic-bounding problem (though it is not the only possible tool for doing so), so in the interest of simplicity I will allow such rewards.

Similar comments hold for other non-licensing positions: (31) does not single out positions between the underlying host and the licensor for rewards in the way in which negative gradient PL penalised just those positions. Knowing which positions are between the underlying host and the licensor again requires access to the underlying configuration. (This was not a problem in the negative version of the constraint, which penalised disharmony and therefore needed only to identify the host(s) and licensor in each candidate.) Furthermore, licensing-driven harmony occasionally goes beyond the licensor and targets an entire word, a phenomenon Walker (2011) calls maximal licensing. This occurs in Eastern Andalusian, for example; see (8a). Whereas Walker develops a separate PL formalism for maximal licensing, (31) offers the possibility of capturing it with the same formalism that motivates standard PL effects. See Kaplan (2017) for analyses, including one for Eastern Andalusian, that demonstrate the advantages of rewarding positions besides the ones between the underlying host and the licensor. In the interest of keeping the current work focused, I do not explore the issue in detail here, though §4.4.2 briefly sketches the proposal in Kaplan (2017) for dealing with the possibility of harmony beyond the licensor.

Positive gradient PL captures the informal generalisations about licensing-based processes. Harmony on the licensor is most important, and nothing is rewarded in its absence. But PL-driven phenomena often exhibit harmony outside the licensor, even occasionally on positions that are not between the trigger and the licensor, and (31) motivates this too. More specifically, as Walker (2011) explains, the functional role of PL is to improve an element's perceptual salience. As she observes, there are (at least) two basic strategies for doing this: put the element in a perceptually prominent position – the licensor – or put it in more than one position, as in maximal licensing. (31) unites these strategies.

Returning to the issues at hand, positive PL cannot block epenthesis, as shown in (33a). In contrast with (29), candidates (ii) and (iv) perform identically with respect to License, because unharmonised positions are irrelevant to PL's assessment. PL also no longer motivates deletion, as shown in (33b); in fact, deletion of potential harmonisers removes a possible locus for a reward from PL.

(33) a.

/ˈtɾeβtol-es/	License([−ATR],$\acute{\sigma}$) 4	Ident 3	NoCoda 2	Dep 1	$\mathcal{H}$
i. ˈtɾeβtolɛ		−1	−1		−5
ii. ˈtɾɛβtolɛ	2	−2	−1		0
iii. ˈtɾeβotolɛ		−1		−1	−4
☞ iv. ˈtɾɛβotolɛ	2	−2		−1	1

b.

/ˈordeni/	License([+high]$_{\text{post-tonic}}$,$\acute{\sigma}$) 3	Ident 2	Max 1	$\mathcal{H}$
i. ˈordeni				0
☞ ii. ˈurdini	3	−2		5
iii. ˈurdni	2	−1	−1	3

As Kimper (2011) shows, positive constraints suffer from an 'infinite goodness' problem, whereby epenthesis of reward-generating elements (in this case harmonising vowels) occurs unchecked, ad infinitum. Some mechanism is required to head this possibility off, and Kimper's solution – the only one I am aware of – is to adopt a serial framework, so that epenthesis occurs on a step before the operation that earns rewards (e.g. harmony): with no motivation for epenthesis absent harmony, that step is impossible, and the derivation cannot move down the infinite-epenthesis path. The same move is required here: the candidate *[ˈurdi̲ini], with the epenthetic vowel underlined, would have a score of 8 if added to (33b). As long as *w*(Dep) < 3, PL motivates unending epenthesis. But if the derivation must first pass through the stage [ˈurdV̲ini], with an epenthetic but unharmonised vowel (see McCarthy 2008 for gradual epenthesis/deletion), epenthesis does not help: this candidate receives the same rewards and penalties as the actual output, plus a penalty from Dep.

Because the infinite-goodness problem strays from the distance-based pathologies at issue in this paper, I leave a full serial implementation of PL for future work. (The outline of a serial derivation from §3.2 remains pertinent for positive gradient PL, however.) In the remainder of the paper I give parallel analyses, and note points at which serialism has substantive consequences.

We can now almost complete our summary table. The results from the foregoing discussion are added in Table IV (the ‘infinite goodness’ cell).

	categorical		gradient	
	negative	positive	negative	positive
parallel	pathologies		no identity licensing	infinite goodness
serial	pathologies		pathologies	correct prediction

Table IV

Consequences of the combinations of various theoretical primitives (version 4).

All that remain are the cells for positive categorical PL. It is easy to see that these configurations, too, are pathological. Positive categorical PL is identical to (10), except that it assigns +1 to a licensed feature instead of −1 to an unlicensed one. This does not substantively affect the asymmetric trade-off we began with. Harmony of any sort earns exactly +1 as long as it includes the licensor, and the cost of this is potentially many Ident violations (for parallelism) or *Skip(V) violations (for serialism). As distance increases, those violations will overwhelm License’s +1. Both gradience and positivity are inadequate without the other.

The completed summary is given in Table V. Only positive gradient PL in a serial framework avoids both the distance-based pathologies and issues such as the infinite-goodness problem.

The following section supports the entry in the bottom right cell of Table V by showing that positive gradient PL produces both indirect and identity licensing without the distance-based pathologies.

4.3 The licensing patterns revisited

The formalism in (31) produces indirect and identity licensing, and it can also model languages with no licensing-based harmony. *Duplicate is

	categorical		gradient	
	negative	positive	negative	positive
parallel	pathologies	pathologies	no identity licensing	infinite goodness
serial	pathologies	pathologies	pathologies	correct prediction

Table V

Consequences of the combinations of various theoretical primitives (version 5).

unnecessary, because, as we saw above, PL now provides the incentive for positions between the trigger and licensor to harmonise. *DUPLICATE is therefore not included here or in subsequent sections; I tentatively conclude that it can be replaced by the current version of PL and, where necessary, *SKIP(V) (e.g. when particular vowels block harmony; see §4.4.1). This section explores the conditions under which indirect and identity licensing are produced, and it demonstrates that the theory is no longer pathological in the way traditional PL is.

Functions for the relevant candidates' harmony scores are provided in (34), where w_L and w_I are the weights of LICENSE and IDENT respectively.

(34)

/ˈee$_0$-i/	LICENSE w_L	IDENT w_I	$\mathcal{H}$
a. ˈee$_0$-i			0
b. ˈie$_0$-i	2	-1	$2w_L - w_I$
c. ˈii$_0$-i	$n+1$	$-n$	$(n+1)w_L - (n \times w_I)$

The faithful form's score is always 0. Consequently, it can only be beaten by a candidate with a positive score: the reward from PL must exceed the penalty from IDENT. For any candidate with x positions that undergo assimilation, the penalty from IDENT is $-w_I \times x$, and LICENSE's reward is $w_L(x+1)$. A positive score is possible, then, only if $w_I \times x < w_L(x+1)$. In the simplest case, identity licensing, $x = 1$, and this candidate therefore has a positive score if $w_I < 2w_L$.

Indirect licensing, where $x = n$, receives additional penalties and rewards compared to identity licensing. Each harmonising intervening position receives +1 from LICENSE and −1 from IDENT. Indirect licensing therefore wins if (i) $w_I < 2w_L$, which permits harmony in the licensor, and (ii) $w_I < w_L$, which allows the intervening positions to harmonise. Since the latter entails the former, indirect licensing merely requires $w_I < w_L$. (Identity licensing, in turn, requires (i) and the absence of (ii).)

The weights that generate each pattern are summarised in (35). As usual, ties are ignored.

(35) a. Indirect licensing: $w_I < w_L$
b. Identity licensing: $w_L < w_I < 2w_L$
c. No harmony: $w_I > 2w_L$

To describe these conditions in other terms, as long as (35a) holds, assimilation of licensing and non-licensing positions is advantageous. To prevent spreading to the intervening positions, the weight of IDENT must therefore be greater than the weight of LICENSE. But this arrangement does not necessarily preclude all assimilation: LICENSE's reward for spreading to the licensor is effectively doubled in comparison to its reward for spreading to other positions, because it assigns +1 for the licensor and +1 again for the underlying host. (Here we see the advantage of permitting PL to reward underlying hosts.) Consequently, as long as the weight of IDENT is not more than twice the weight of LICENSE, the latter will motivate assimilation of the licensor (cf. (35b)). Finally, if the weight of IDENT is more than twice the weight of LICENSE, harmony fails (cf. (35c)).

The combinations of weights that yield indirect licensing, identity licensing and no harmony are shown in Fig. 4.

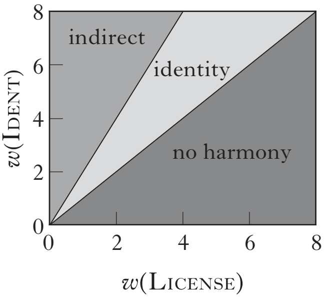


Figure 4

The outcomes for various combinations of weights for IDENT and LICENSE.

The tableaux in (36) give example weights that favour each candidate type, and Fig. 5 gives the corresponding graphs. In all three graphs, there is a single winner for $n > 1$, confirming that distance-based pathologies have been vanquished.[6] Changing constraint weights affects only (i) whether identity licensing outperforms the faithful form, and (ii) whether indirect licensing's function has a positive slope.

---

[6] Though $n = 1$ is less interesting for the reasons given above, in all cases the outcome at $n = 1$ is compatible with the outcome at $n > 1$: either the faithful candidate wins, or indirect and identity licensing tie at $n = 1$ – and for our purposes indirect and identity licensing are indistinguishable when $n = 1$ – and one of those options wins for $n > 1$.

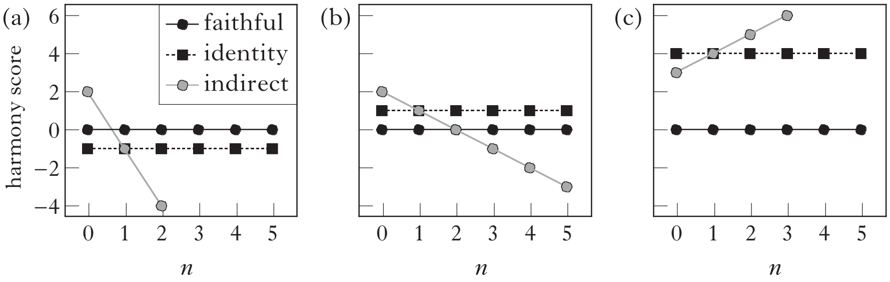


Figure 5

Harmony scores for the candidates and weights in (a) (36a), (b) (36b) and (c) (36c).

(36) a. *No licensing*

/ˈee$_0$-i/	License 2	Ident 5	$\mathcal{H}$
☞ a. ˈee$_0$-i			0
b. ˈie$_0$-i	2	−1	−1
c. ˈii$_0$-i	$n+1$	$-n$	$2(n+1)-5n$

b. *Identity licensing*

/ˈee$_0$-i/	License 2	Ident 3	$\mathcal{H}$
a. ˈee$_0$-i			0
☞ b. ˈie$_0$-i	2	−1	1
c. ˈii$_0$-i	$n+1$	$-n$	$2(n+1)-3n$

c. *Indirect licensing*

/ˈee$_0$-i/	License 3	Ident 2	$\mathcal{H}$
a. ˈee$_0$-i			0
b. ˈie$_0$-i	2	−1	4
☞ c. ˈii$_0$-i	$n+1$	$-n$	$3(n+1)-2n$

(35) holds only for what we might call the canonical case: a single non-licensing position hosts the feature restricted by LICENSE, and the feature spreads to or is copied in the licensor. If multiple positions host the restricted feature underlyingly, the weighting conditions might change somewhat. This and other non-canonical situations are discussed in § 5.

The next section provides further support for positive, distance-sensitive PL by putting it to use in analyses of Central Veneto and Eastern Andalusian.

4.4 Analyses of Central Veneto and Eastern Andalusian

4.4.1 *Central Veneto.* Recall from (1) that in the variety spoken in Central Veneto, post-tonic [+high] triggers raising of the stressed vowel, with concomitant raising of vowels between the trigger and target. The analysis follows straightforwardly from (35): $w_I < w_L$ yields the right outcome, as shown in (37).

(37)

/ˈordeni/	License([+high]$_{\text{post-tonic}}$,σ́) 3	Ident[high] 2	$\mathcal{H}$
a. ˈordeni			0
b. ˈurdeni	2	−1	4
☞ c. ˈurdini	3	−2	5

Metaphony is blocked by /ɛ ɔ a/. They do not raise when stressed, as in (38a), and, when they appear between the trigger and licensor, both resist raising themselves and prevent the stressed vowel from raising, as in (38b).

(38)

a.	ˈvɛʧi		‘old man (masc pl)’
	ˈtɔki		‘piece (masc pl)’
	ˈgati		‘cat (masc pl)’
b.	laˈ(v)orava	laˈ(v)oravi	‘work (1/2sg impf ind)’
c.	ˈpɛrsego	ˈpɛrsegi	‘peach (masc sg/pl)’
	ˈangolo	ˈangoli	‘angle (masc sg/pl)’

Furthermore, when the stressed vowel may not raise, the intervening vowels do not raise either, as shown in (38c). This was the evidence provided in §4.2 for making PL’s reward for intervening vowels contingent upon harmony in the licensor.

Spreading [+high] to /ɛ ɔ a/, all of which are [−ATR], would create [ɪ ʊ], which are unattested in the language. Consequently, their resistance to raising reflects two highly weighted constraints (Walker 2011): *ɪ,ʊ (which blocks /ɪ ʊ/) and Ident[ATR] (which prevents raising to [i u]). Adding these to the analysis prevents harmony of any sort with stressed /ɛ ɔ a/, as shown in (39).[7]

(39)

/ˈangol-i/	Ident[ATR] 8	*ɪ,ʊ 8	License([+high]$_{\text{post-tonic}}$,σ́) 3	Ident[high] 2	$\mathcal{H}$
☞ a. ˈangoli					0
b. ˈunguli	−1		3	−2	−3
c. ˈʊnguli		−1	3	−2	−3
d. ˈanguli				−1	−2

---

[7] It should be apparent from this tableau that a sufficient reward from License can overcome either of the new constraints. This is dealt with below in (41), once all the necessary constraints are in place.

In subsequent tableaux I omit *ɪ,ʊ and candidates with [ɪ ʊ], leaving candidates like *[ˈunguli] to represent illicit raising.

Currently, the analysis predicts identity licensing when the intervening vowel cannot raise (*[laˈ(v)uravi]), as shown in (40). License motivates assimilation of the stressed syllable, and Ident[ATR] requires the intervening vowel's faithfulness.

(40)

/laˈ(v)oravi/	Ident[ATR] 8	License([+high]$_{\text{post-tonic}}$,σ́) 3	Ident[high] 2	ℋ
☜ a. laˈ(v)oravi				0
b. laˈ(v)uruvi	−1	3	−2	−3
👎 c. laˈ(v)uravi		2	−1	4

*Skip[a] – a member of Kimper's (2012) *Skip(V) family – prevents this outcome by penalising gapped configurations in which [a] is skipped, as in (41), while *Skip[ɛ] and *Skip[ɔ] deal with the other non-raisers.

(41)

/laˈ(v)oravi/	Id[ATR] 8	*Skip[a] 5	Lic([+high]$_{\text{post-tonic}}$,σ́) 3	Id[high] 2	ℋ
☞ a. laˈ(v)oravi					0
b. laˈ(v)uruvi	−1		3	−2	−3
c. laˈ(v)uravi		−1	2	−1	−1

Two concerns remain: with sufficiently many intervening vowels, (i) the reward from License can overcome the penalty for raising a stressed /ɛ ɔ a/ (/ˈaee-i/ → *[ˈuii-i]), and (ii) License and *Skip[a] can conspire to force indirect licensing that includes raising of intervening non-raisers like [a] (/ˈeea-i/ → *[ˈiiu-i]). Both patterns arise because the reward from License (and in (ii) avoiding the penalty from *Skip[a]) more than compensates for the single violation of Ident[ATR].

Both issues are addressed by serialism, which is independently necessary to avoid the infinite-goodness problem (see §4.2). Recall from §3.2 that only one vowel harmonises per step, beginning with the stressed vowel. This solves both problems, because the anti-harmony pressure from the non-raisers only has to contend with harmony on the stressed syllable, not harmony everywhere. Harmony of stressed /ɛ ɔ a/ earns +2 from License, so as long as the penalty for that harmony from Ident[ATR] and Ident[high] counteracts this reward, stressed /ɛ ɔ a/ cannot raise, regardless of the intervening vowels. Thus we need w(Ident [ATR]) + w(Ident[high]) > 2w(License) – a condition met in the tableaux above. (42) illustrates the point with a hypothetical form. With only one vowel able to harmonise, the +2 from License cannot overcome faithfulness. The candidate [ˈungulutututut-i], the pathological output, would have a score of +1, but it is an illicit candidate at this step.

(42)

/ˈangolotototot-i/	Id[ATR] 8	*Skip[a] 5	License 3	Id[high] 2	$\mathcal{H}$
☞ a. ˈangolotototot-i					0
b. ˈungolotototot-i	−1		2	−1	−4

As for problem (ii), intervening non-raisers cannot be forced to harmonise as long as the penalty from *Skip[a] and Ident[high] is greater than the reward for harmonising the stressed vowel. That is, the first step in the derivation /ˈeea-i/ → [ˈiiu-i] passes through [ˈiea-i], but when $w(*\textsc{Skip}[a]) + w(\textsc{Ident}[high]) > 2w(\textsc{License})$, this step is impossible. This condition, too, is met in the preceding tableaux. The competition between candidates (a) and (c) in (41) illustrates the point: this comparison represents the first step of the derivation just described, and no number of intervening licit raisers can save candidate (c), because only the stressed vowel may raise at this step.

4.4.2 *Eastern Andalusian.* The salient difference between the licensing systems of Central Veneto and Eastern Andalusian is that Eastern Andalusian's intervening vowels optionally harmonise, as shown in (7) above (though if one does, they all do).

Weights that conform to (35b) produce Eastern Andalusian's identity licensing, as (43) shows. Candidate (43d), with only one harmonised intervening vowel, is collectively harmonically bounded by candidates (b) and (c), giving the all-or-nothing nature of harmony on these vowels. In Eastern Andalusian, the harmony trigger is itself unfaithful (being the product of /s/-aspiration), so each candidate's score is reduced by 3 (i.e. one Ident violation) compared to analogous candidates from previous sections.

(43)

/ˈkometel-os/	Ident[ATR] 3	License([−ATR],σ́) 2	$\mathcal{H}$
a. ˈkometelɔ	−1		−3
☞ b. ˈkɔmetelɔ	−2	2	−2
c. ˈkɔmɛtɛlɔ	−4	4	−4
d. ˈkɔmetɛlɔ	−3	3	−3

High vowels undergo word-final laxing, but they resist harmony, as shown in (8b) above. This fact is simple to accommodate with *ɪ,ʊ, from the analysis of Central Veneto, and Max[−ATR]. In (44), *ɪ,ʊ prevents the stressed vowel from harmonising, and Max[−ATR] ensures that the final vowel does not escape laxing; recall that in the view of Jiménez & Lloret (2007) and Lloret & Jiménez (2009), laxing reflects preservation of a deleted /s/'s [spread glottis] feature.

(44)

/ˈkɾisis/	Max[−ATR] 6	Ident[ATR] 3	License([−ATR],σ́) 2	*ɪ,ʊ 2	ℋ
☞ a. ˈkɾisɪ		−1		−1	−5
b. ˈkɾɪsɪ		−2	2	−2	−6
c. ˈkɾisi	−1				−6

For completeness, (45) shows how the weights can be adjusted to produce Eastern Andalusian's indirect-licensing variants by adopting weights conforming to (35a). For discussion of manipulating weights to produce variation in HG, see Jesney (2007), Pater *et al.* (2007) and Hayes (2017). I leave it for future work to flesh out a rigorous analysis of this optionality (for example, with the reduction in Ident's weight, the weight of *ɪ,ʊ must be increased to retain the result in (44)), but this gives an indication of what such an analysis might look like.

(45)

/ˈkometel-os/	License([−ATR],σ́) 2	Ident[ATR] 1	ℋ
a. ˈkometelɔ		−1	−1
b. ˈkɔmetelɔ	2	−2	2
☞ c. ˈkɔmɛtɛlɔ	4	−4	4
d. ˈkɔmetɛlɔ	3	−3	3

This analysis is extended in Kaplan (2017) to account for pre-tonic harmony by capitalising on the fact that License rewards harmony on positions beyond the licensor. The gist of the analysis is given in (46): by varying the relationship between License and Ident[ATR]$_{\text{pre-tonic}}$, pre-tonic harmony can be produced (46a) or suppressed (46b).

(46) a.

/reˈkohelos/	License 4	Ident[ATR] 2	Ident[ATR]$_{\text{pre-tonic}}$ 1	ℋ
i. reˈkohelɔ		−1		−2
ii. reˈkɔhelɔ	2	−2		4
iii. reˈkɔhɛlɔ	3	−3		6
iv. rɛˈkɔhelɔ	3	−3	−1	5
☞ v. rɛˈkɔhɛlɔ	4	−4	−1	7

b.

/reˈkohelos/	License 2	Ident[ATR]$_{\text{pre-tonic}}$ 2	Ident[ATR] 1	ℋ
i. reˈkohelɔ			−1	−1
ii. reˈkɔhelɔ	2		−2	2
☞ iii. reˈkɔhɛlɔ	3		−3	3
iv. rɛˈkɔhɛlɔ	4	−1	−4	2

Candidate (46a.iv), with pre-tonic but not post-tonic harmony, is harmonically bounded, so the analysis correctly predicts that pre-tonic harmony occurs only when post-tonic harmony also appears.

5 Remaining issues

Throughout, I have assumed the canonical case described in §4.3: an unlicensed feature appears in exactly one position in the input, and it must find its way to the licensor. Under such conditions, the requirements in (35) hold. But the balance between License and faithfulness changes under different circumstances.

In the canonical case, identity licensing (e.g. /ˈeee-i/ → [ˈiee-i]) earns +2 from License and −1 from Ident; indirect licensing with the same input (yielding [ˈiii-i]) earns +4 from License and −3 from Ident. But if a second position hosts the relevant feature underlyingly – as in /ˈeei-i/ – License's reward for identity licensing ([ˈiei-i]) increases to +3, and Ident's penalty for indirect licensing decreases to −2. The consequence is a greater incentive for harmony and the possibility that a grammar could treat the inputs /ˈeee-i/ and /ˈeei-i/ differently. The extent of this possibility and whether corrective steps are needed to avoid it remain to be seen. I know of no data from Central Veneto with multiple post-tonic high vowels, and in Eastern Andalusian, /s/-aspiration is the only source of [−ATR], so only one trigger – the final vowel – is possible.

The analyses developed above also considered only overwrite systems, in which unlicensed features trigger assimilation in the licensor, as opposed to preservation systems, in which either unlicensed features are eradicated or the licensor triggers harmony in non-licensing positions (see Kaplan 2015). I have not tried to account for preservation here, nor have I ruled out candidates that show preservation alternatives to the overwrite found in Central Veneto and Eastern Andalusian. The issue is one of directionality: which position must assimilate to the other? The decision is often left to positional faithfulness (Beckman 1999), but Walker (2011) identifies a number of other ways of dictating directionality.

Imagine a language in which positional faithfulness favours preservation, so unlicensed features are eradicated, comparable to Classical Mongolian (see §3.1): /ˈe-i/ → [ˈe-e]. But if the reward for overwrite is large enough (i.e. if there are many intervening positions to spread to) License can compel a change in directionality, producing overwrite: /ˈeee-i/ → [ˈiii-i]. This is reminiscent of 'majority rules' effects (Baković 2000); serialism may rule out this pattern by pitting faithfulness against harmony just in the licensor, not against harmony in many positions.

Walker's (2011) third licensing pattern, direct licensing, which was mentioned briefly in §3.1, is unaccounted for in the current system. Walker uses CrispEdge constraints (Itô & Mester 1999, Walker 2001, Kawahara 2008) to confine elements to the licensor in Esimbi's vowel-height transfer system and related patterns: when CrispEdge and PL outrank faithfulness, the restricted element surfaces exclusively in the licensor. It is easy to imagine that the accumulation of faithfulness violations can prevent satisfaction of CrispEdge, and this possibility requires careful attention. CrispEdge also has consequences for the weighting conditions in (35), as it is violated by both indirect and identity licensing.

Finally, because positive gradient PL rewards harmony of non-licensors, it invites the possibility of runaway harmony that targets as many non-licensors as possible: assimilation may continue beyond the licensor or extend in the opposite direction from the licensor. In Kaplan (2017) I argue that this has its advantages – most relevantly, it permits an account of pre-tonic harmony in Eastern Andalusian – but the possibilities need fuller exploration. Walker (2011) considers other unbounded PL-driven harmony systems that both lend support to the version of PL developed above and provide good testing grounds.

6 Alternatives

I have argued that the remedy for the distance-based pathologies rests on a reformulation of PL. This section considers other approaches to the problem. Each falls short.

6.1 Infinite weights

Prince & Smolensky (1993) observe that OT's strict domination amounts to an HG grammar that permits infinite weights: the ranking License $\gg$ Ident is reproduced in HG if $w(\text{License}) = \infty$. There are two reasons this does not provide a solution to the pathological interaction of these constraints. First, unless we decree that License may have a weight of either infinity or zero, but nothing in between, we still predict grammars with pathological weights. Second, with a weight of infinity, License cannot be outweighed by any constraint, at least not in an obvious way,[8] and therefore no constraint can mitigate its effects. We have already seen that this is inaccurate: Central Veneto requires Ident[ATR] to outweigh PL.

6.2 *Skip(V)

*Skip(V) is an appealing alternative to distance-sensitive PL, because it can penalise the failure of intervening positions to harmonise in proportion to the number of such positions that are present, much as the version of PL in (23) does. It therefore potentially counters Ident's escalating penalties.

*Skip(V) can indeed eliminate the Identity at a Distance pathology, as (47) shows: under these weights, identity licensing has a score of $-2(n-1)-1$, which is worse than indirect licensing's $-n$ for any $n > 1$. If

---

[8] An anonymous reviewer notes that one kind of infinity might outweigh another (e.g. $\aleph_0$ *vs.* $\aleph_1$). But, as the reviewer acknowledges, this significantly complicates the computation of harmony. This is not a path to go down without more compelling motivation.

indirect licensing beats identity licensing at short distances, it does so at long distances, too.

(47)

/ˈee$_0$-i/	License 5	*Skip(V) 2	Ident 1	$\mathcal{H}$
☞ (n>5) a. ˈee$_0$-i	−1			−5
b. ˈie$_0$-i		$-(n-1)$	−1	$-2(n-1)-1$
☞ (n<5) c. ˈii$_0$-i			$-n$	$-n$

But the No Distant Licensing pathology remains. Neither the faithful candidate nor the indirect-licensing candidate violates *Skip(V) (as (47) shows), so this constraint cannot stop Ident violations from overwhelming License. This simply reinforces the argument made above: distance-based pathologies are symptomatic of the relationship between PL and faithfulness, and therefore the solution is to be found in those constraints.

To make things worse, *Skip(V) extends the No Distant Licensing pathology to include identity licensing. (Originally, the No Distant Licensing pathology only affected indirect licensing, because only there do Ident violations accumulate.) As the distance between the trigger and target increases, the score for identity licensing becomes worse, because of the escalating penalty from *Skip(V). If weights different from those in (47) are selected that favour identity licensing at short distances, the faithful form, with its static score, will eventually outperform identity licensing. *Skip(V) is useful in HG-based PL analyses, as the analysis of Central Veneto shows, but it does not correct the pathologies at hand.

6.3 Changing faithfulness

The distance-based pathologies arise because of an asymmetry between PL and faithfulness; if we eliminate the asymmetry, we eliminate the pathologies. In previous sections I showed how this is achieved by transforming PL into a gradient constraint. Here I consider the other possibility: changing faithfulness so that it does not assign violations in proportion to the number of positions that harmonise. I present two possibilities: categorical Ident (§6.3.1) and Max[F] constraints (§6.3.2). The first leads to unwanted predictions in other domains, and the second simply does not eliminate the pathologies.

6.3.1 *Categorical faithfulness.* To avoid escalating Ident violations, we might redefine Ident so that it assigns a single violation no matter how many segments are unfaithful. Ident assigns −1 to /ee$_0$-i/ → [ie$_0$-i] and to /ee$_0$-i/ → [ii$_0$-i], for example. This is a pyrrhic victory, though, because categorical Ident introduces vast problems. For example, it invites counting effects whereby the accumulation of markedness violations can compel the violation of higher-weighted faithfulness. (48)

shows a language in which one mid vowel is permitted in a word, because w(Ident[high]) > w(*Mid), but when a second mid vowel appears, both vowels raise. This pattern emerges because raising any number of vowels incurs a single Ident violation, so while raising one vowel is disadvantageous, raising two means avoiding two *Mid violations at the cost of one Ident violation. By adjusting the weights, the maximum number of mid vowels can be arbitrarily established.

(48) a.	/bed/	Ident[high] 3	*Mid 2	$\mathcal{H}$
	☞ i. bed		−1	−2
	ii. bid	−1		−3
b.	/bede/			
	i. bede		−2	−4
	ii. bide	−1	−1	−5
	☞ iii. bidi	−1		−3

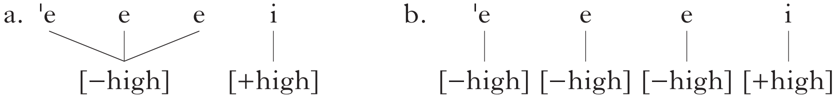
Other pathologies stem from this core issue. Segments that are otherwise compelled to be faithful (because w(Faith) > w(Markedness)) can take a ‘free ride’ if another segment must be unfaithful. In (49), even though Ident[voi] has several times the weight of *VoiObs, the fact that *VoiObs$_{coda}$ triggers final devoicing in (49b) means that the onset can be devoiced for free: it does not incur a second violation of Ident[voi].

(49) a.	/ba/	*VoiObs$_{coda}$ 8	Ident[voi] 7	*VoiObs 1	$\mathcal{H}$
	☞ i. ba			−1	−1
	ii. pa		−1		−7
b.	/bag/				
	i. bag	−1		−2	−10
	ii. bak		−1	−1	−8
	☞ iii. pak		−1		−7

In eliminating the asymmetric trade-off between PL and faithfulness, categorical faithfulness merely introduces new asymmetric trade-offs in other areas. I therefore conclude that it is not a viable option.

6.3.2 Max[F]. For some features, and in certain circumstances, the Max family of constraints is preferable to the Ident family (Lombardi 1998). This is an appealing place to turn to for the problem at hand, for the following reason. If we assume that in the schematic form /ˈeee-i/ the first three vowels share a single [−high] feature, as in (50a), then spreading the final vowel’s [+high] to each vowel violates Max[high] just once (because only a single [−high] is deleted), in contrast with the three IDENT [high] violations that such spreading entails.

(50)



With MAX[high], then, it seems that faithfulness violations do not escalate with the distance between the trigger and target. But this is of course simply a consequence of the feature-sharing representation in (50a). If we instead assume that each vowel hosts a unique [−high] feature, as in (b), then escalation re-emerges: spreading to all the non-high vowels incurs three MAX[high] violations. By richness of the base, (50b) cannot be excluded, so MAX[F] constraints do not escape distance-based pathologies. As with *SKIP(V), MAX[F] is valuable, but does not solve these particular problems.

Similar remarks hold for other kinds of faithfulness constraints, such as existential faithfulness constraints (Struijke 2000). They behave non-pathologically in particular configurations, but the problematic asymmetric trade-off re-emerges elsewhere. However, see O'Hara (2016) for evidence that MAX[F] has some utility in weeding out pathologies in non-PL-driven harmony.

7 Discussion and conclusion

Gradient constraints in OT, especially alignment constraints, are notoriously controversial, because they 'assign violation scores in an unusual manner' (Pater 2009: 1019). Alignment is computationally more demanding or more powerful than other constraints (Eisner 1997a, b, Potts & Pullum 2002, McCarthy 2003), and, as we saw above, gradience invites pathologies.

On the basis of an asymmetric trade-off very much like the one examined here, Legendre *et al.* (2006) criticise gradient constraints from the point of view of HG. Their example involves the interaction of STRESSHEAVY (which demands stress on heavy syllables) and MAINSTRESS-R (which gradiently penalises main stresses that are not right-aligned in a word). Because greater misalignment incurs greater penalties from MAINSTRESS-R, this constraint can force stress off of heavy syllables (violating higher-weighted STRESSHEAVY exactly once) if the heavy syllable is too far from the right edge of the word. Thus heavy syllables are stressed only when they fall within an arbitrary window at the right edge of a word. After arguing that such systems are unattested, Legendre *et al.* conclude that '*grammars can't count*' (2006: 344; emphasis original).

In their example, eradicating the asymmetric trade-off by recasting the analysis in categorical terms is possible, as McCarthy (2003) demonstrates. But this solution is unavailable to PL. I argued in §6.3.1 that changing the constraint that behaves gradiently in PL systems – faithfulness – has unattractive outcomes. We can learn two things from this result. First, otherwise categorical constraints like IDENT can appear to be gradient from the point of view of constraints like PL. Perhaps there is no sharp division between gradient and categorical constraints after all; there are simply well-behaved and unruly constraints of both types. In this sense, the traditional formulation of PL is *too* categorical.

Second, gradient constraints can sometimes have advantages over their categorical counterparts. A gradient version of PL establishes the right kind of balance with faithfulness where the categorical version could not. This, combined with the proposals by McCarthy (2010) (involving serialism) and Kimper (2011) (involving positivity and serialism) for reining in gradient constraints' most pernicious properties, implies that gradient constraints are well-formed members of CON after all – at least in HG.

It has occasionally been observed that 'an OT analysis of any linguistic pattern can be translated into an HG one' (Pater *et al.* 2007: 2). Demonstrations to this effect are found in Prince & Smolensky (1993) and Legendre *et al.* (2006), for example. The difficulty of devising a theory of PL for HG that replicates the typology achieved in Walker's (2011) OT framework highlights the limits of this claim – or, more accurately, the conditions under which it is true and the entities that it holds for. Metaphony-like patterns arise in OT under the ranking LICENSE ≫ FAITH, and for any finite set of input–output pairs showing that sort of harmony, there are weights for PL and faithfulness that replicate the OT result, even under the traditional form of PL given in (10). We need only make the weight of LICENSE greater than the product of the weight of FAITH and the maximum distance that harmony crosses in the extant data. (This claim holds more broadly: any finite set of outputs produced by OT can be replicated with the same constraints in HG; see Pater 2009 and references therein.) In that sense, the observation quoted at the beginning of this paragraph is true, but only if we take the 'linguistic pattern' to be that finite set of input–output pairs – the extensional pattern. If we instead take it to be the intensional linguistic system (that we believe is) exemplified by the finite set of input–output pairs – say, [+high] invariably spreads from a final vowel to the stressed vowel – then the traditional PL analysis in OT cannot be replicated in HG. To put it differently, the ranking LICENSE ≫ FAITH predicts that unbounded spreading will reach the licensor; no HG analysis with the same constraints can make the same prediction (in the absence of infinite weights). All of this is to say that we must bear in mind the distinction between the data at hand and the linguistic system we believe it reveals. A single set of constraints may model the former in both OT and HG, but (as Pater 2009, for example, observes) capturing the latter can necessitate different instantiations of CON in the two frameworks.

I have argued here that PL is one of these loci of divergence between OT and HG. Recasting PL in gradient and positive terms places it on firmer ground in HG, and though it now looks quite different from its OT counterpart, this is the first step toward a theory of PL that is as successful in HG as it has been in OT.

REFERENCES

Baković, Eric (2000). *Harmony, dominance and control*. PhD dissertation, Rutgers University.

Beckman, Jill N. (1999). *Positional faithfulness: an Optimality Theoretic treatment of phonological asymmetries*. New York: Garland.

Cole, Jennifer & Charles Kisseberth (1994). An optimal domains theory of harmony. *Studies in the Linguistic Sciences* **24**. 101–114.

Crosswhite, Katherine M. (2001). *Vowel reduction in Optimality Theory*. New York & London: Routledge.

Eisner, Jason (1997a). FootForm decomposed: using primitive constraints in OT. *MIT Working Papers in Linguistics* **31**. 115–143.

Eisner, Jason (1997b). What constraints should OT allow? Handout from paper presented at the 71st Annual Meeting of the Linguistic Society of America, Chicago. Available as ROA-204 from the Rutgers Optimality Archive.

Goldsmith, John A. (1989). Licensing, inalterability, and harmonic rule application. *CLS* **25:1**. 145–156.

Hayes, Bruce (2017). Varieties of Noisy Harmony Grammar. In Karen Jesney, Charlie O'Hara, Caitlin Smith & Rachel Walker (eds.) *Proceedings of the 2016 Meeting on Phonology*. http://dx.doi.org/10.3765/amp.v4i0.3997.

Hayes, Bruce & Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* **23**. 59–104.

Hyman, Larry M. (1988). Underspecification and vowel height transfer in Esimbi. *Phonology* **5**. 255–273.

Itô, Junko (1986). *Syllable theory in prosodic phonology*. PhD dissertation, University of Massachusetts, Amherst.

Itô, Junko & Armin Mester (1999). Realignment. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*. Cambridge: Cambridge University Press. 188–217.

Jesney, Karen (2007). The locus of variation in weighted constraint grammars. Poster presented at the workshop 'Variation, gradience and frequency in phonology', Stanford University. Available (February 2018) at http://www-bcf.usc.edu/~jesney/Jesney2007Variation.pdf.

Jesney, Karen (2011). Licensing in multiple contexts: an argument for Harmonic Grammar. *CLS* **45:1**. 287–301.

Jiménez, Jesús & Maria-Rosa Lloret (2007). Andalusian vowel harmony: weak triggers and perceptibility. Paper presented at the workshop 'Harmony in the languages of the Mediterranean', 4th Old World Conference in Phonology, Rhodes. Handout available as ROA-901 from the Rutgers Optimality Archive.

Kaplan, Aaron (2008). *Noniterativity is an emergent property of grammar*. PhD dissertation, University of California, Santa Cruz.

Kaplan, Aaron (2011). Harmonic improvement without candidate chains in Chamorro. *LI* **42**. 631–650.

Kaplan, Aaron (2015). Maximal prominence and a theory of possible licensors. *NLLT* **33**. 1235–1270.

Kaplan, Aaron (2017). Overshoot in positional licensing. Paper presented at the Western Conference on Linguistics (WECOL) 2017, Boise State University. Handout available (February 2018) at https://linguistics.utah.edu/faculty/aaron_kaplan/Overshoot%20in%20Positional%20Licensing.pdf.

Kawahara, Shigeto (2008). On the proper treatment of non-crisp-edges. In Mutsuko Endo Hudson, Sun-Ah Jun, Peter Sells, Patricia M. Clancy & Shoichi Iwasaki (eds.) *Japanese/Korean linguistics*. Vol. 13. Stanford: CSLI. 55–67.

Kimper, Wendell (2011). *Competing triggers: transparency and opacity in vowel harmony*. PhD dissertation, University of Massachusetts Amherst.

Kimper, Wendell (2012). Harmony is myopic: reply to Walker 2010. *LI* **43**. 301–309.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness: an application. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum. 884–891.

Legendre, Géraldine, Antonella Sorace & Paul Smolensky (2006). The Optimality Theory–Harmonic Grammar connection. In Paul Smolensky & Géraldine Legendre (eds.) *The harmonic mind: from neural computation to optimality-theoretic grammar*. Vol. 2: *Linguistic and philosophical implications*. Cambridge, Mass.: MIT Press. 339–402.

Lloret, Maria-Rosa & Jesús Jiménez (2009). Un análisis *óptimo* de la armonía vocálica del andaluz. *Verba* **36**. 293–325.

Lombardi, Linda (1994). *Laryngeal features and laryngeal neutralization*. New York: Garland.

Lombardi, Linda (1998). Evidence for MaxFeature constraints from Japanese. *University of Maryland Working Papers in Linguistics* **7**. 41–62. Available as ROA-247 from the Rutgers Optimality Archive.

McCarthy, John J. (2003). OT constraints are categorical. *Phonology* **20**. 75–138.

McCarthy, John J. (2004). Headed spans and autosegmental spreading. Ms, University of Massachusetts, Amherst. Available as ROA-685 from the Rutgers Optimality Archive.

McCarthy, John J. (2006). Restraint of analysis. In Eric Baković, Junko Itô & John J. McCarthy (eds.) *Wondering at the natural fecundity of things: essays in honor of Alan Prince*. Santa Cruz: Linguistics Research Center. 213–239.

McCarthy, John J. (2008). The gradual path to cluster simplification. *Phonology* **25**. 271–319.

McCarthy, John J. (2010). Autosegmental spreading in Optimality Theory. In John A. Goldsmith, Elizabeth Hume & W. Leo Wetzels (eds.) *Tones and features*. Berlin & Boston: De Gruyter Mouton. 195–222.

McCarthy, John J. & Alan Prince (1993). Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

Martin, Andrew T. (2005). *The effects of distance on lexical bias: sibilant harmony in Navajo compounds*. MA thesis, University of California, Los Angeles.

O'Hara, Charlie (2016). Harmony in Harmonic Grammar by reevaluating faithfulness. *NELS* **46:3**. 71–84.

Parker, G. W. (1883). *A concise grammar of the Malagasy language*. London: Trübner.

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

Pater, Joe, Rajesh Bhatt & Chris Potts (2007). Linguistic optimization. Ms, University of Massachusetts, Amherst. Available as ROA-924 from the Rutgers Optimality Archive.

Poppe, Nicholas (1954). *Grammar of written Mongolian*. Wiesbaden: Harrassowitz.

Poppe, Nicholas (1955). *Introduction to Mongolian comparative studies*. Helsinki: Suomalais-Ugrilainen Seura.

Potts, Christopher & Geoffrey K. Pullum (2002). Model theory and the content of OT constraints. *Phonology* **19**. 361–393.

Prince, Alan (2003). Anything goes. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata & Shin-ichi Tanaka (eds.) *A new century of phonology and phonological theory: a Festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday*. Tokyo: Kaitakusha. 66–90.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.

Samek-Lodovici, Vieri & Alan Prince (1999). Optima. Ms, University College London & Rutgers University. Available as ROA-363 from the Rutgers Optimality Archive.

Samek-Lodovici, Vieri & Alan Prince (2005). Fundamental properties of harmonic bounding. Ms, University College London & Rutgers University. Available as ROA-785 from the Rutgers Optimality Archive.

Sanders, Benjamin P. (1998). The Eastern Andalusian vowel system: form and structure. *Rivista di Linguistica* **10**. 109–135.

Stallcup, Kenneth L. (1980a). A brief account of nominal prefixes and vowel harmony in Esimbi. In Luc Bouquiaux (ed.) *L'expansion bantoue*. Vol. 2. Paris: Société d'Études Linguistiques et Anthropologiques de France. 435–441.

Stallcup, Kenneth L. (1980b). Noun classes in Esimbi. In Larry M. Hyman (ed.) *Noun classes in the Grassfields Bantu borderland. Southern California Occasional Papers in Linguistics* **8**. 139–153.

Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John J. McCarthy & Joe Pater (2010). OT-Help 2.0. Software package. http://people.umass.edu/othelp.

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

Struijke, Caro (2000). *Existential faithfulness: a study of reduplicative TETU, feature movement, and dissimilation*. PhD dissertation, University of Maryland, College Park.

Walker, Rachel (2001). Round licensing, harmony, and bisyllabic triggers in Altaic. *NLLT* **19**. 827–878.

Walker, Rachel (2004). Vowel feature licensing at a distance: evidence from northern Spanish language varieties. *WCCFL* **23**. 773–786.

Walker, Rachel (2005). Weak triggers in vowel harmony. *NLLT* **23**. 917–989.

Walker, Rachel (2010). Nonmyopic harmony and the nature of derivations. *LI* **41**. 169–179.

Walker, Rachel (2011). *Vowel patterns in language*. Cambridge: Cambridge University Press.

Wilson, Colin (2003). Analyzing unbounded spreading with constraints: marks, targets, and derivations. Ms, University of California, Los Angeles.

Zoll, Cheryl (1997). Conflicting directionality. *Phonology* **14**. 263–286.

Zoll, Cheryl (1998a). *Parsing below the segment in a constraint-based framework*. Stanford: CSLI.

Zoll, Cheryl (1998b). Positional asymmetries and licensing. Ms, MIT. Available as ROA-282 from the Rutgers Optimality Archive.

Zymet, Jesse (2015). Distance-based decay in long-distance phonological processes. *WCCFL* **32**. 72–81.