Error-Driven Learning in Harmonic Serialism[*]

Anne-Michelle Tessier

University of Alberta & University of Michigan

1 Introduction: learning from errors in a serial constraint-based phonology

Harmonic Serialism (HS) is a constraint-based but derivational view of phonology and its typology, suggested first by Prince and Smolensky (1993/2004: 79) and proposed first in detail in McCarthy (2000, 2007ab). As a blend between serial and parallel phonological grammar, HS has demonstrated some desirable advantages over classic Optimality Theory for capturing attested phonological typologies without overgenerating (examples and arguments in McCarthy 2008ab; Jesney 2008/2012; Pruitt 2010; Kimper 2010; Elfner, to appear.) If HS is to be embraced as a theory of phonology, however, it must be associable with a working theory of learning (see also Staubs and Pater, to appear). This paper is an initial foray into how difficult it is to learn the phonotactics of an HS grammar, and how these difficulties might in principle be overcome.

The main challenge comes from the class of 'hidden rankings', and the problem is sketched here very briefly. HS derivations happen step-by-step, driven by one markedness constraint at time, but an error-driven learner will only see the end result of that derivation. To build a grammar that can drive each step in between, HS sometimes requires a ranking between markedness constraints (M1 >> M2) *which is not revealed by errors.* Without this hidden ranking, the learner will acquire an incorrectly-variable grammar: one which sometimes produces a superset of the target language's attested forms. This paper's proposed solution is to reason not just from the learner's errors but also their *winners* – that is, observed surface forms they have stored – and to compare those winners to a particular class of their failed loser candidates, through an iterative process that will eventually produce errors that evince the hidden ranking. The broader argument is that straight phonotactic learning is rather more complicated in HS than in OT, but HS also makes every input's candidate set much less complicated, and that this trade-off may be a key to successful HS learning. To set the stage, the next two sections

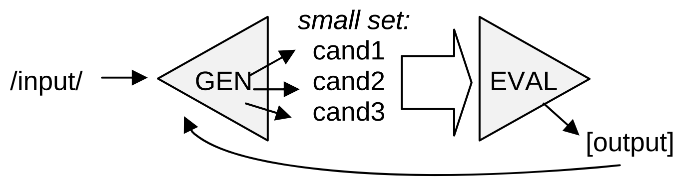
[*] Thanks especially to Karen Jesney for raising the crucial question and improving my insights, and to Michael Becker, Andries Coetzee, Brett Hyde, Emily Elfner, Matt Goldrick, John McCarthy, Joe Pater, Kathryn Pruitt, Robert Staubs, Matt Wolf, the MIT Phonology Circle, particularly Adam Albright, Ed Flemming and Igor Yanovich, the UMichigan PhonDi Group and Karthik Durvasula, and the audience at NELS42, especially Claire Moore-Cantwell. All errors are sadly the author's alone.

provide a quick introduction to the Harmonic Serialism framework and its derivations, and the error-driven approach to learning ordinal constraint rankings which will later illustrate the HS learning problem.

2 The basics of a Harmonic Serialist phonology

The HS architecture is illustrated in figure (1) below. As in OT, a Harmonic Serialist-grammar feeds a single input to a GEN function, which returns a set of candidates, and which together are fed to an EVAL function to produce an optimal output.

1) An HS schematic:



HS uses the same EVAL function as OT: built on a language-specific basis from a basic set of markedness and faithfulness constraints in a total ordering, and winnowing candidates via strict domination down to a single optimum. HS differs fundamentally in two ways, though. First, its GEN function returns a much smaller, finite set of candidates which differ minimally from the input, defined below[1]. Second, the HS framework allows for multiple input→output loops until the ranking converges on a final winner.

To see how candidate sets are chosen and convergence is determined, I now walk through a hypothetical HS derivation, using the central input and constraint set example for this paper. The input is /maki/, from which GEN must first build a candidate set. Guided by McCarthy **(**2007a: 61–62, 77–79**),** I assume GEN returns all those candidates which each violate only one of the 'basic' faithfulness constraints, only once. This 'basic' set might include MAX, DEP and IDENT-FEATURE and perhaps a few others, but in this example I limit GEN to just two faith constraints: IDENT[CONTINUANT] and IDENT[VELAR]. Thus, /maki/'s entire candidate set has only three members: [maki], [maxi] (violating only IDENT[CONT]) and [maci] (violating only IDENT[VELAR]).[2] As this example is very small, I provide here the scehmatic results of applying GEN to all of four crucial inputs in this paper:

2) *Exhaustive range of GEN for four input arguments*

a) GEN(/...aki/) = {...aki, ...axi, ...aci} * ...**açi**
b) GEN(/...aci/) = {...aci, ...açi, ...aki} * ...**axi**
c) GEN(/...axi/) = {...axi, ...aki, ...açi} * ...**aci**
d) GEN(/...açi/) = {...açi, ...aci, ...axi} * ...**aki**

This CON also contains two markedness constraints, which drive fairly simply and well-attested featural processes. The first is labeled in tableaus as *V-STOP-V, a lenition constraint prohibiting intervocalic stops (as in Tiberian Hebrew: Gurevich 2004; Kaplan,

[1] The precise nature of this candidate set remains a crucial subject of HS research: see McCarthy (2010).
[2] Under the assumption that IDENT[VELAR] only militates against velar vs. palatal featural changes.

2010). The second is labeled *XI, prohibiting velar consonants followed by high vowels (as seen in Greek fricative allophony, Kazazis, 1972). The constraint ranking used throughout this illustrative derivation is in (3) below:

3) *V-STOP-V >> *XI >> IDENT[CONT], IDENT[VELAR]

In the first step of parsing /maki/, the ranking chooses among (3a)'s three possible candidates, and *V-STOP-V causes /k/ to spirantize:

4) HS derivation step one: stop spirantization

/maki/	*V-STOP-V	*XI	IDENT [CONT]	IDENT [VELAR]
maki	*!			
☞ maxi		*	*	
maci	*!			*

The HS grammar now proceeds by assessing whether the winner is the fully faithful candidate: in (4) it has mapped *un*faithfully from /k/ to [x], and this means that the optimal candidate from step one's derivation is taken as the input to step two. Input /maxi/ is given to GEN, which provides again a very limited candidate set as seen in (2c). Note that [maci] is no longer a candidate: it is too distant from this input, because /x/→[c] is unfaithful *both* to place and continuancy.

5) HS derivation step two: velar fricative fronting

/maxi/	*V-STOP-V	*XI	IDENT [CONT]	IDENT [VELAR]
maxi		*!		
☞ maçi				*
maki	*!		*	

At step 2, the velar fricative in /maxi/ has fronted to avoid violating *XI – and since this meant another unfaithful mapping, the grammar again submits this new output as input to a third step. From [maçi] GEN builds the set of minimally-different candidates in (2d) – but this time, the fully-faithful candidate is in fact the winner:

6) HS derivation step three: convergence

/maçi/	*V-STOP-V	*XI	IDENT [CONT]	IDENT [VELAR]
☞ maçi				
maxi		*!		*
maci	*!		*	

Tableau (6)'s fully-faithful mapping signals that the grammar has converged on a winner: [maçi] has been chosen as the final optimum.

This walk-through shows how serial and gradual HS is (and isn't). While each EVAL step is taken in parallel, GEN's cautious restrictiveness ensures that processes driven by markedness constraints are undertaken one at a time. In this grammar: the pressure of *V-STOP-V first pushes /k/ to lenite to [x]; this in turn causes a violation of lower-ranked *XI, which at the second step pushes /xi/ to [çi]. In the terms of Kiparsky (1968), the ranking *V-STOP-V >> *XI has put the process of intervocalic spirantization and velar fronting into a *feeding* relationship, and the ultimate topic of this paper is how an HS learner can acquire such feeding relations.

3 The basics of ERC-driven phonological learning

The general learning approach in this paper is adopted from considerable previous work on the learnability of phonotactics using a classic OT grammar: particularly Tesar and Smolensky (1998, 2000), Hayes (2004) and Prince and Tesar (2004). This particular learner begins with a maximally restrictive grammar hypothesis of all markedness constraints above all faithfulness constraints {M} >> {F}, and aims to learn a ranking that will faithfully reproduce all and only the surface forms that it observes (that is, it knows nothing yet of morphology or unfaithful mappings due to allomorphy; see esp. Tesar, 2004; Hayes, 2004). In (7) I provide this example's initial state ranking, which adds one new markedness constraint prohibiting palatal fricatives (*C):

7) *Initial state: {M} >> {F}*
{*V-STOP-V, *XI, *C} >> {IDENT[CONT], IDENT[VELAR]}

The learner begins by using its initial state ranking to parse observed target forms, which I will assume happens each time by resolving any partial orderings into a randomly-chosen consistent total ordering. When an observed form as input is *not* reproduced faithfully, the learner identifies this mapping as an error.

Suppose that in our hypothetical target language, the restriction on [xi] sequences enforced in (5) forms part of a velar/palatal fricative allophony (again, as in modern Greek): palatals are found just before high front vowels, and velar fricatives elsewhere. In this case, errors will be made on those occasions when (7) is translated into a total ordering like (8) below, where *C >> XI. Any such ranking will incorrectly ban palatal fricatives across the board, so observed forms like /maçi/ will be mapped unfaithfully:

8) A possible initial state error on palatal fricatives

/maçi/	*V-STOP-V	*C	*XI	IDENT [CONT]	IDENT [VELAR]
maçi		*!			
☞ maxi			*!		*
maci	*!			*	

Having made an error, the learner then converts stores it as in (9): an ERC vector (Prince, 2002). This ERC indicates which constraints prefer the observed form [maçi], which in the target language is the intended Winner (noted with a W), and which prefer the current grammar's optimum [maxi], which in the target grammar is a Loser (L); constraints that prefer or disprefer both options equally are noted with an e:

9) An ERC vector, built from (8)

input	*winner ~ loser*	*V-STOP-V	*XI	*C	IDENT[CONT]	IDENT[VELAR]
/maçi/	[maçi] ~ [maxi]	**e**	**W**	**L**	**e**	**W**

Learning proceeds by analyzing a set of ERCs and building a new ranking, using an algorithm which both resolves the errors and otherwise remains as restrictive as possible. To summarize: the algorithms (i) resolve errors by ranking *some* Winner-preferring constraint above *all* Loser-preferring constraints, and (ii) ensure restrictiveness by ranking as many markedness constraints above faithfulness constraints as possible. Given (9), these two principles determine the new ranking *XI >> *C to prevent de-palatalization before high vowels (resolving the ERC), but otherwise keep to the initial state ranking as much as possible, producing the new ranking in (10):[3]

10) {*V-STOP-V, *XI} >> *C >> {IDENT-CONT, IDENT-VELAR}

3.1 Using ERCs to learn in Harmonic Serialism

To use the ERC-driven phonotactic learner from the previous section with a Harmonic Serialist grammar, one practical problem must be addressed. HS derivations make one change at a time, so each *step* is easily translated into an ERC, but the entire mapping may contain several steps. Recalling section 2's crucial example: what is the correct ERC to store if an error was made like /maki/ → maxi → [maçi]? In classic OT, the winner vs. loser *maki* ~ *maçi* can be directly compared, but HS cannot assess multiple unfaithful mappings at a time; there is no tableau in which maki and maçi are compared.

Among the various possible technical solutions, I propose here a fairly simple and conservative one.[4] When the HS learner makes an error with multiple unfaithful mappings, they simple store an ERC from the derivation's *first step*. Schematically: if a learner is attempting to reproduce an observed form /A/ → [A], and instead they produce an error via the derivation /A/ → B → [C], the resulting ERC would be built just from the /A/ → [B]. This is illustrated in (11) below:

11) A schematic ERC vector, built from the first step of HS derivation /A/ →B → [C]

input	*winner ~ loser*	*A	*B	*C	IDENT A/B	IDENT B/C
/A/	[A] ~ [B]	L	W	e	W	e

This first-step error will teach the learner something small but crucial: namely that in this target language, *B >> *A. Later on the learner may need to also determine that the /B/ → [C] mapping, which was their initial error's *second* step, is also incorrect, but that can wait for more overt evidence.[5]

[3] For explicit characterizations of this class of ranking algorithms, see Prince and Smolensky (1993/2004), Tesar and Smolensky (1998, 2000), Hayes (2004), Prince and Tesar (2004), Tessier (2007, 2009).
[4] Thanks to Karen Jesney for generously providing this suggestion and discussion of alternatives.
[5] For further discussion of how learning from first-step ERCs can slowly but accurately bring about a target ranking see Tessier (2012).

4 The problem: learning hidden rankings in HS

With an ERC-driven Harmonic Serialism learner in place, this section lays out the problem of learning HS phonotactics. Building from the previous sections, the full target language to be learned is illustrated in (12): stops become fricatives intervocalically (a), velar fricatives become palatal before [i] and remain velar elsewhere (b), and when a velar stop precedes a high vowel, both processes apply (c). Recalling that the phonotactic learner only observes target forms and does not posit unfaithful inputs, the full set of learning evidence available to cause errors and re-ranking is in (13):

12) a) *Intervocalic spirantization*
/mat/ → [mat] /mati/ → [maθi]
/map/ →[map] /mapi/ → [mafi]

b) *Velar/palatal fricative allophony*
/maxi/ ↘ /max/ → [max]
/maçi/ → [maçi] /maç/ ↗

c) *Spirantization feeds palatalization*
/mak/ →[mak] /maki/ → maxi → [maçi]

13) *Observed surface forms from* (12)
[mat] [maθi] [map] [mafi] [mak] [max] [maçi]

The ranking in (14) represents this papers's target: i.e., the HS grammar that produces (12)'s language. As will be seen directly, all three markedness constraints must be ranked in this order in HS to produce (12)'s patterns.

14) *Target ranking necessary to derive* (12), *via faithful mappings of* (13)
*V-STOP-V >> *XI >> *C >> {IDENT-CONT, IDENT-VELAR}

4.1 Failing to learn the HS hidden ranking

The challenge for the HS phonotactic learner is encoding the feeding relationship in (12c). At the initial state (recall tableau 7) all three markedness constraints are unranked at the top of the hierarchy. The examples in (7)-(9) showed how this initial ranking will cause errors that teach the learner (12b)'s allophony, leading to the ranking repeated in (15) below. Notice also that this ranking enforces the restriction in (12a), via the ranking algorithm's inherent {M >> F} bias:

15) *Result of ERC-driven phonotactic learning, given* (13)
{*V-STOP-V, *XI} >> *C >> {IDENT-CONT, IDENT-VELAR}

The problem, however, is that *no* phonotactic errors will demonstrate the facts of (12c). The available data in (13) makes this clear: there are no surface forms that violate *XI or *V-STOP-V, since no intervocalic segments are stops, and no velar fricatives precede high vowels. If no observed forms cause a markedness constraint to be assigned a L in an ERC vector (as was the case for *C in 9) then the learner's bias will keep that markedness constraint undominated in the ranking.

Thus, the ranking of *V-STOP-V and *XI with respect to each other is *hidden* from the HS learner. In section 2, their correct ranking was required to cause feeding between spirantization and palatalization of input velars in a hypothetical form like/maki/. But without such hypothetical forms, the learner has acquired (15)'s indeterminate ranking between the two markedness constraints – and this represents a failure of phonotactic learning. When tested on a hypothetical input /maki/, the ranking in (15) must first settle into a total ordering, with either *V-STOP-V or *XI highest ranked. If the correct ranking as in (14), the grammar will correctly derive /maki/ → maxi → [maçi] as seen already in section 2. But when the chosen ranking is as shown in (16) below, the grammar incorrectly predicts that /maki/ maps faithfully:

16) Restrictiveness failure: one ranking from (15) that maps /maki/ → *[maki]

/maki/	***XI**	***V-STOP-V**	*C	IDENT[CONT]	IDENT[VELAR]
☞ maki		*			
maxi	*!			*	
~~maçi~~	*not in the candidate set for* /maki/				

The flaw in (16)'s ranking, given the HS framework, is that there is no gradual way for /maki/ to be optimized. The highest ranked markedness constraint *XI blocks /maki/ from taking the necessary first step towards [maçi], and so the derivation converges with the fully faithful candidate. As a result, the language created by (15) vacillates between mapping /maki/ to [maçi] and *[maki]; the learner has incorrectly acquired a grammar in which intervocalic stops are generally prohibited – except, sometimes, in context [Vki].

Before moving on, it is worth recalling why this learning problem is specific to the HS framework. The need for a *gradual* path of markedness improvement, from one input to the next, is only necessary because HS insists on incremental changes. In OT, the tableau in (16) for input /maki/ would include not just similar outputs but the full set proviuded an OT GEN, including the intended winner [maçi]. Since /maki/ → [maçi] can be chosen there directly, both *XI and *V-STOP-V can be satisfied in one fell swoop, so the grammar in (15) remains restrictive despite its unranked top stratum.

The upshot of this section is that errors on observed forms alone will not reveal all the necessary phonotactic rankings of a Harmonic Serialist grammar. The next section proposes a novel way to find the necessary evidence to learn hidden HS rankings.

5 The proposal: learning hidden rankings from Winners

The overall approach to acquiring HS hidden rankings here is to break down learning into three stages. First, the learner uses ERCs made via errors on observed forms to construct necessary and unhidden rankings: for example, making errors on *[maçi] as in (8), from which our learner discovered that *XI >> *C. Later, the learner examines their ranking for lurking hidden rankings among markedness constraints, and uses the proposal below to decide whether any of them are crucial for ensuring restrictiveness.[6]

[6] I set aside here the crucial third step of setting up unfaithful inputs to account for surface allomorphy, as discussed in OT by Kager, 1999; Jarosz, 2006; Tesar and Prince, 2007 *interalia*.

The proposal's precise goal is an algorithm which, when given the constraint set of (7) and observed forms in (13), finds the hidden ranking in (14): *V-STOP-V >> *XI. This will require building an ERC which assigns a W to V-STOP-V and an L to*XI – such an ERC is shown in (17)[7]:

17) An ERC vector that illustrates the hidden ranking

winner ~ loser	*V-STOP-V	*C	*XI	IDENT[CONT]	IDENT[VELAR]
[maxi] ~ [maki]	**W**		**L**		

The biggest difficulty with (17) is that it requires an *ungrammatical winner*, which by definition will not be observed among the observed surface forms during phonotactic learning. To see that unlenited /k/ is disprefered to lenited [x] in this context, we must hypothesize both of these structures to begin with, even though the surface language in (13) provides evidence of neither. Thus, the learner needs a principled way to hypothesize inputs like /maki/ and winners like [maxi]. For ease of reference, I will call this process Hidden Ranking Discovery (HRD).

The HRD process starts with the ranking at the end state of phonotactic learning shown in (15) and repeated below in (18):

18) *Result of ERC-driven phonotactic learning, given* (13)
{*V-STOP-V, *XI} >> *C >> {IDENT[CONT], IDENT[VELAR]}

In HRD, the will treat each pair of equally-ranked markedness constraints M1 and M2 in turn as an unanswered question: *does restrictiveness require a ranking among {M1, M2}*?

5.1 HRD Step One: Examining Failed Candidates and Constructing ERCs

The necessary beginning is to look beyond errors and ERCs – and to examine *winners*, that is, those known observed forms that the current grammar can reproduce faithfully.[8]

The learner first takes the set of observed target forms (the Winners) and feed each one to GEN, each time returning a set of output candidates; these I will call the Intermediate Forms (IFs). Since the current grammar maps each Winner onto its own fully faithful candidate, the set of IFs for a particular winner also represents that winner's set of failed losers, as they each fare worse than a fully faithful parse. The learner now constructs a set of ERCs, using each IF as input,[9] and treating the target form as winner

[7] I have not given this ERC an input, because only the violation profiles of markedness constraints are important. In other words: using either /maki/ or /maxi/ as an input would be equally effective.

[8] To be sure, this means that the learner must be storing observed forms as they go, in a way that their phonology can access – on which skeptics are encouraged to see Becker and Tessier (2011).

[9] As in footnote 5: this input is actually irrelevant to this section's entire process, because all the Ws and Ls that will cause re-ranking are markedness constraints. However, the ERCs produced in these steps will be combined with normal phonotactic errors later on, which do contain violations of faithfulness constraints, so leaving the faithfulness comparisons undefined here will make later steps complicated. Thus, I make explicit how inputs are chosen for constructed ERCs, but nothing crucial hinges on them.

and an IF as loser. After building this ERC set using every IF in turn, the learner narrows down all of these winner~loser pairs to the following crucial set: *those pairs in which M1 prefers the winner (i.e. assigns more violations to the loser) and M2 prefers them equally (i.e. assigns them equal violations), and whose loser is not represented among the known observed forms.*

This process is illustrated in (19) and (20) below, using as input a potential real word of the target language, [baçi]. The tableau in (19) shows the faithful winner compared with the results of GEN(baçi), its two IF losers (recall (2d)). The two resulting ERCs both meet the criteria above to be included in the crucial set: compared to the winner (19b), *V-STOP-V prefers the winner in (19a), while *XI is silent; in (19c) *XI prefers the winner and *V-STOP-V is silent. Finally, neither IF has been observed in the target (cf. (13), which contains neither [baxi] nor [baci].)

19) A faithful mapping and two losers

/ baçi/	*V-STOP-V	*XI	*C	IDENT [CONT]	IDENT [VELAR]
a) baci	*!			*	
☞ baçi			*		
b) baxi		*!			*

These crucial ERCs are shown in (20), comparing (19b~a) and (19b~c) respectively:

20) 'Constructed ERCs' based on (19)'s *losers*

input	*winner~loser*	*V-STOP-V	*XI	*C	IDENT[CONT]	IDENT[VELAR]
a) / baci/	baçi ~ baci	W			L	
b) / baxi/	baçi ~ baxi		W	L		L

These constructed ERCs do not yet teach the learner anything new compared to the ranking in (15) – this is not surprising, since (15) already preferred the winner [baçi] over both losers. Instead, Step one's ERCs simply create evidence for existing rankings of M1 and M2: (20a) demonstrates why *V-STOP-V must outrank faithfulness, and (20b) provides additional evidence that *XI >> *C.

What Step One is really about is identifying useful Intermediate Forms: a set of hypothetical inputs whose unfaithful mappings are relevant to the relative ranking of M1 or M2 in isolation. For example, (20b)'s ERC turns the learner's attention to the question '*if* /baxi/ *were given as input to the grammar, why would faithful *[baxi] lose out to target [baçi]*?' – and the answer is that the winner's violation of *C is preferable to the loser's violation of *XI, and therefore that *XI >> *C.[10]

5.2 HRD Step Two: Building the complement of constructed ERCs

In Step 2, the focus moves one step further back from the target language. To do so, it re-applies most of Step 1 – but based this time not on target words, but rather Intermediate Forms. Step 1 showed that one of the two markedness constraints, call it M1, is

[10] The learner's inherent bias to keep M >> F means that IDENT[VELAR] will remain low-ranked.

responsible for mapping each IF onto a target output: e.g. in (20b), *XI drove /baxi/ to [ba**ç**i]. Step 2's question is whether M2 can drive a similar but *prior* unfaithful mapping.

To answer this question, the learner takes each IF as mapped unfaithfully by a constraint M1, and applies GEN to yield its small candidate set; these I will recall Remote Forms or RFs (since they are further away from a target output than the IFs). The learner again builds a new set of ERCs: using each Remote Form as input, paired with its IF as winner and a faithful RF candidate as loser. Among the full set of these new ERCs, the learner again finds a crucial set as follows: *those pairs in which the other relevant constraint, call it M2, prefers the IF winner (i.e. assigns more violations to the remote form) and M1 prefers the RF loser (i.e. assigns more violations to the intermediate form) and whose loser is not represented among the known observed forms.*

Back in our ongoing example, one Intermediate Form in Step 1 was /baci/ which violated *V-STOP-V as its M1. The set of two Remote Forms that GEN returns for /baci/ (cf. 2) are used to build two potential ERCs, as Step 2 describes, provided in (21). The violation profiles iimmediately reveal that neither fits in the crucial set – not only does M2, *XI, not choose any winners, in fact *no* constraints choose a winner, meaning that this constraint set cannot be ranked to prefer these winners to losers. (Note also that the (21a) is also out of the running because its loser is *not* unattested.)

21) Potential IF ERCs created with GEN(baci) – *no crucial ERCs*

input	*winner~loser*	*V-STOP-V	*XI	*C	IDENT[CONT]	IDENT[VELAR]
a) /baçi/	baci ~ baçi	L	e	e	L	e
b) /baki/	baci ~ baki	e	e	e	e	L

Thus the learner finds no evidence from this IF for any hidden rankings.

The other IF from Step 1 was /baxi/, which violated *XI as its M1. The /baxi/ set of Remote Forms created by GEN are again used to build two potential ERCs, as in (22):

22) Potential IF ERCs created with GEN(baxi) – *one crucial ERC!*

input	*winner~loser*	*V-STOP-V	*XI	*ç	IDENT[CONT]	IDENT[VELAR]
a) /baki/	**baxi ~ baki**	**W**	**L**	**e**	**L**	**e**
b) /baçi/	baxi ~ baçi	e	W	L	e	W

Unlike the previous case, this ERC table *does* provide a winner~loser pair that meets step 2's crucial set criteria. The constraint M2 for this IF was *V-STOP-V, and (22a)'s pair *baxi ~ baki* has exactly the right violation profile: M2 assigns a W, and M1 (*XI) assigns an L, plus its loser [baki] is ungrammatical in the target. And this error is exactly the form spelled out at the beginning of this section in (17) as the crucial evidence of the hidden ranking. Now the learner just has to use this ERC to learn.

5.3 HRD Step Three: Learning from the Accumulated ERCs

The final step is to simply build the union of the crucial ERC sets found in steps 1 and 2, along with the normal ERCs learned from observed forms during phonotactic learning, and feed them all together to the re-ranking algorithm. For our example, this full set of

ERCs is compiled in (23) below (each row is labeled with its location earlier in the paper for reference). The first error was made organically by the initial state ranking; the next two were created in step 1 and the final error was created in step 2.

23) **Full** set of ERCS for phonotactic + hidden ranking learning

input	*winner~loser*	*V-STOP-V	*XI	*ç	ID[CONT]	ID[VELAR]
(9) /maci/	maçi ~ maxi		**W**	**L**		W
(20a) / baci/	baçi ~ baci	**W**			L	
(20b) / baxi/	baçi ~ baxi		**W**	**L**		L
(22a) /baki/	baxi ~ baki	**W**	**L**		L	

When given (23). the re-ranking algorithm will finally learn the correct restrictive grammar, including the hidden ranking, which I now demonstrate briefly (for formalisms, see Prince and Tesar, 2004; Hayes, 2004). The algorithm first installs in Stratum 1 those markedness constraints that *prefer no Losers*: i.e. those constraints, which from these errors, appear undominated in the target (see 24). That resolves error (20a) and (22a), since both losers are now ruled out in Stratum 1, leaving two errors unexplained (see 25).

24) Stratum 1: ***V-STOP-V**

25) Remaining set of ERCS to resolve after stratum 1

input	*winner~loser*	*V-STOP-V	*XI	*ç	ID[CONT]	ID[VELAR]
(9) /maci/	maçi ~ maxi		**W**	**L**		W
(20b) / baxi/	baçi ~ baxi		**W**	**L**		L

In Stratum 2 the learner again install all markedness constraints that prefer no Losers, in (26) – *and now the hidden ranking has been created.* Finally, since Stratum 2's constraint resolves the two remaining errors, the learner installs the rest of their constraints via inherent biases (i.e. {M} >> {F}). The total ranking, then, is in (27):

26) Stratum 1+2: ***V-STOP-V >> *XI**

27) Final ranking**: *V-STOP-V >> *XI >> *C>> {IDENT-CONT, IDENT-VELAR}**

The ranking in (27) represents success: the learner has learned a feeding relationship between spirantization (via *V-STOP-V) and fricative allophony (controlled by *XI >> *C), even in the absence of overt phonotactic evidence.

6. Insufficient Discussion and Interim Conclusions

Before concluding, this section touches very briefly on two of the important questions about this proposal and its implications. The first is the range of possible outcomes from HRD, and its chances of either failing to find or unnecessarily imposing hidden rankings. The second concerns the role of alternations in the acquisition of HS hidden rankings.

The first two steps of the Hidden Ranking Discovery process simply results in a set of ERCs. The re-ranking algorithm that uses those ERCs in step 3 is well-supported by previous literature cited above; so long as the ERCs being fed to the algorithm are legitimate, we can be confident that the result will not include incorrect rankings (at least any that cannot later be overcome from positive data.) But what makes an ERC 'legitimate'? By calling an ERC legitimate, I mean that its loser really is suboptimal compared to its winner in the target language. This property of step 1 and 2s constructed ERCs is maintained by the last clause describing both steps' crucial ERC set: *the loser must not be observed in the language.* In step 1 the winner *is* observed, so if the grammar maps loser to winner, the ranking will be target-appropriate for phonotactic purposes. In step 2 the winner is by definition *not* part of the language, being an IF – but if the loser is not in the language either, there is no harm mapping the RF loser to IF winner, since step 1 has already created an ERC that will map IF loser to observed winner.

As for the possible outcomes: if HRD finds an ERC in Step 2 like (23d), then a new ranking between markedness constraints will be built into the ranking in Step 3; if it doesn't, nothing about the phonotactic grammar will change. The effects of the process are thus quite minimal – though note that it may need to be applied many times, if there are many crucially unranked markedness constraints within a stratum {M1, M2, M3, M4}, and the implications of ordering the search for hidden rankings, e.g. comparing {M1, M2} before {M2, M3} if all three must be crucial ranked, are not yet known.

Another crucial question is how the acquisition of alternations, after phonotactic learning and HRD are complete, interacts with the learner's knowledge of hidden rankings. Clearly, some hidden rankings are brought out of hiding via alternations: in our example, a target language like (13) which had a root [mak] and a suffix [-i] would provide the morphologically-aware learned with a paradigm [mak] ~ [maçi], creating an error or errors that underscore the need to map /k/ somehow to [ç], albeit via multiple steps. (The acquisition of alternations within a Harmonic Serialist framework also remains open for future research.) The current HRD procedure, however, will be necessary just in case such allomorphy does not happen to occur in the lexicon – cf. the related argument made by McCarthy (1998) regarding a learning bias for high-ranking OO-faith to ensure restrictiveness without alternations.

As stated in the introduction, the first crucial point of this paper is that hidden rankings in HS pose a serious problem for the phonotactic learner, because observed forms and the errors they cause will not reveal all the crucial rankings among markedness constraints. However, the proposal's other crucial point is that HS provides a promising method of finding *unobserved* forms to consider in learning, via its restricted GEN. Because every input's candidate set is both finite and highly structured, a learner may be able to reason cautiously away from the forms it knows to construct informative learning data about forms it has not yet heard – without sacrificing restrictiveness, or overall efficiency.

References

Becker, Michael and Anne-Michelle Tessier. 2011. Trajectories of faithfulness in child-specific phonology. Phonology 28(2): 163-196.

Emily Elfner. To appear. Stress-epenthesis interactions in Harmonic Serialism. In Harmonic Grammar and Harmonic Serialism, ed. John J. McCarthy and Joe Pater. London: Equinox Press.

Gurevich, Naomi. 2004. Lenition and Contrast: The Functional Consequences of Certain Phonologically Conditioned Sound Changes. Outstanding Dissertations in Linguistics. New York: Garland Publishing.

Hayes, Bruce. 2004. Phonological Acquisition in Optimality Theory: the early stages. In Fixing Priorities: Constraints in Phonological Acquisition, ed. René Kager, Joe Pater and Wim Zonneveld, 158-203. Cambridge, U.K.: Cambridge University Press.

Jarosz, Gaja. 2006. Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory. Doctoral dissertation, John Hopkins University, Baltimore MD.

Jesney, Karen 2008/2012. Positional Faithfulness, non-locality, and the Harmonic Serialism solution. Proceedings of NELS 39, ed. Suzi Lima, Kevin Mullins and Brian Smith, 401-412. Amherst, MA: GLSA.

Kazazis, Kostas. 1972. Some remarks on palatalization in Modern Greek. Research on Language & Social Interaction 5(2): 330–332.

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

Kaplan, Abby. 2010. Phonology shaped by phonetics: The case of intervocalic lenition. Doctoral dissertation, UC Santa Cruz.

Kimper, Wendell. 2011. Locality and globality in phonological variation. Natural Language and LinguisticTheory 29(2): 423-465.

Kiparsky, Paul. 1968. Linguistic universals and linguistic change. In Universals in Linguistic Theory, ed. Emmon Bach and Robert Harms, 170-202. New York: Holt, Rinehart & Winston.

McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. In Proceedings of CLS 35, Vol. II: The Panels, ed. M. C. Gruber, D. Higgins, K. Olson, and T. Wysocki, 123-150. Chicago Linguistics Society, Chicago IL.

McCarthy, John J. 2000. Harmonic Serialism and Parallelism. In Proceedings of NELS 30, ed. M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim, 501-524. Amherst, MA: GLSA.

McCarthy, John J. 2007a. Hidden generalizations: phonological opacity in Optimality Theory. London: Equinox.

McCarthy, John J. 2007b. Restraint of analysis. In Freedom of Analysis, ed. Sylvia Blaho, Patrik Bye, and Martin Krämer, 203-231. Berlin: Mouton de Gruyter.

McCarthy, John J. 2008a. The gradual path to cluster simplification. Phonology 25:271-319.

McCarthy, John J. 2008b. The serial interaction of stress and syncope. Natural Language & Linguistic Theory (26(3): 499-546.

McCarthy, John J. 2010. Studying Gen. Journal of the Phonetic Society of Japan 13(2): 3-12.

Prince, Alan and Paul Smolensky. 1993/2004. Optimality Theory: Constraint Interaction in Generative Grammar. Oxford, UK: Blackwell.

Prince, Alan and Bruce Tesar 2004. Learning Phonotactic Distributions. Fixing Priorities: Constraints in Phonological Acquisition, ed. René Kager, Joe Pater and Wim Zonneveld, 245-291. Cambridge, UK: Cambridge University Press.

Pruitt, Kathryn. 2010. Serialism and locality in constraint-based metrical parsing. Phonology 27(3): 481-526.
Smolensky, Paul 1996. On the comprehension/production dilemma in child language. Linguistic Inquiry 27: 720–31.
Staubs, Robert and Joe Pater. to appear. Learning serial constraint-based grammars. In Harmonic Grammar and Harmonic Serialism, ed. John. J. McCarthy and Joe Pater, London: Equinox.
Tesar, Bruce. 2004. Contrast analysis in phonological learning. Ms., Dept of Linguistics and RuCCS, Rutgers University.
Tesar, Bruce, and Alan Prince (2007). Using phonotactics to learn phonological alternations. In Proceedings of CLS 39, Vol. II: The Panels, ed. A. Franklin, J. Cihlar, D. Kaiser, I. Kimbara, 209-237. Chicago Linguistics Society: Chicago, IL.
Tessier, Anne-Michelle. 2009. Frequency of Violation and Constraint-based Phonological Learning. Lingua 119(1): 6-38.
Tessier, Anne-Michelle. 2012. Learning in Harmonic Serialism and the Necessity of a Rich Base. Ms., University of Alberta and University of Michigan.

Department of Linguistics
4-22 Assiniboia Hall
University of Alberta
Edmonton, AB T6G 2E7
CANADA

amtessier@ualberta.ca