The nature of regressions in the acquisition of phonological grammars*

Anne-Michelle Tessier
University of Alberta

1 Introduction

Under one typical view, children's acquisition of their L1 phonological grammar is understood as a gradual progression from an initial universal state towards a language-specific one, in which learners respond to mismatches between their outputs and the ambient language (i.e. their 'errors') by changing their grammars incrementally to better approximate the target (i.e. 'resolving' their errors). One challenging problem for this view are the many reports of 'U-shaped development' in which production temporarily regresses, diverging further from the target rather than drawing closer: see e.g. Menn (1971) *et seq*; Macken (1980); Vihman and Velleman (1989), (2002); Bleile and Tomblin (1991); Bernhardt and Stemberger (1998); Stemberger, Bernhardt and Johnson (1999); Becker and Tessier (2011). To what extent do such regressions cast doubt on the view of phonological acquisition as a gradual process of grammatical error resolution?

Based on existing and novel analyses of longitudinal data, this paper argues that phonological regressions should <u>not</u> be captured directly within the normal workings of children's error-driven mechanisms for grammar learning. Section 2 defines the crucial, problematic type of U-shaped development – *grammatical backtracking* – and claims that grammatical backtracking is restricted to child-specific processes, suggesting an exceptional treatment of these regressions via child-specific constraints that are induced over the course of learning (in the spirit of Becker and Tessier, 2011; see also Inkelas and Rose, 2008). Beyond this limited grammatical backtracking (and other types of regressions which are argued not to be grammatical in nature, see 2.4) section 3 identifies the kind of regression that seems plausible but is nonetheless apparently unattested: one in which markedness constraints flip-flop over time, so that improvement on one marked structure entails regression on another. With this initial empirical base, section 4 demonstrates that an error-driven OT-like learner which stores its errors and imposes certain persistent biases can in fact easily regress in the unattested way. Section 5 discusses how OT's grammatical parallelism is in part responsible for creating the unattested regression pattern, and how a serial constraint-based grammar like Harmonic Serialism (McCarthy 2007 *et seq*) avoids this regression.

2 U-shaped phonology, including grammatical backtracking

2.1 *Grammatical backtracking* This section's goal is to establish the pattern which will be referred to as grammatical backtracking. In this scenario, a process emerges in a child's phonology, often not seen at the earliest production stages, which initially affects a subset of lexical items and then spreads through a phonologically-defined part of the lexicon. It is not known how common or typical this scenario is in natural phonological development – as it may only be noticed within a fairly close and longitudinal study – but it is clearly attested in many diary studies.

One well-documented example is found in Daniel's nasal harmony patterns, as documented in Menn (1971) and summarized in (1) below. At the first stages, nasal consonants caused regressive nasalization of very few (and only labial) stops, and therefore did not affect items like *stone* and *down*; by stage 5,

* Thanks in a not particularly relevant order to Tara McAllister Byun, Eric Bakovic, Michael Becker, Karen Jesney, Joe Pater, Nazarré Merchant and John McCarthy, and to audiences at Michigan State University and Phonology 2013 UMass Amherst, for comments, questions, clarification and discussion. Thanks also to Shelley Velleman and Yvan Rose for asking questions in 2006 and 2012 respectively that eventually got me working on this topic (again).

Proceedings of *Phonology 2013*
Completed December 20, 2013

however, these words had regressed to follow a more general harmony pattern, affecting initial labial and coronal stops alike:

(1) Daniel's U-shaped development of nasal harmony (Menn, 1971)

stages 1-2		*stage 3*		*stages 4-5*	
broom	[mum]	train	[ŋaɪn]	blimp	[mɪmp]
bump	[bʌmp]	prune	[nun], [mum]	bump, jump	[mʌmp]
stone	[don]	stone	[don]	stone	[non]
down	[daʊn]	down	[daʊn]	down	[naʊn]

A second example comes from a systematic study of regression in one child's phonology (Bleile and Tomblin, 1991) which tracked this child's velar fronting over many months. In the initial stages most syllable-initial velar stops mapped to coronals (as seen frequently in child phonologies, e.g. Chiat, 1983; Inkelas and Rose, 2007), but a small set of lexical items resisted fronting:

(2) K's U-shaped development of velar fronting (Bleile and Tomblin, 1991)

Velar Fronting		*Initial Exceptions*		*Regression*
All stages		Stages 1-6		Stages 7-22
candy	[tændi]	clown	[kaʊn]	[taʊn]
		okay	[o'keɪ]	[o'teɪ]
		cookie	['kʊki]	['tʊki]

Bleile and Tomblin (1991) cites seven words – *clown, okay, cookie, kitty cat, (ice) cream* and *Gumbi* – which during stages 1-6 were fronted in only 3 out of 19 total tokens. During stages 7-22, however, they had clearly regressed, as 58/61 tokens of these seven words now showed fronting.

Another example comes from Becker and Tessier (2011)'s quantitative study of place harmony in the corpus of one child Trevor (Compton and Streeter, 1977; Pater, 1977). While most of Trevor's harmony patterns initially affected the majority of potential targets (e.g. between 1;0-1;8 most /TVK/[1] words harmonized to [KVK] or [TVT]), the /KVT/ lexical items showed a U-shaped development in their harmony, whereby harmony in /KVT/ targets became *more* likely between 1;0 and roughly 2;3 (see Becker and Tessier, 2011 for details.)

2.2 *A definition of grammatical backtracking* How problematic are the previous section's data for the 'gradual progression' view of phonological acquisition outlined in the introduction? One point is that these regressions appear to target a phonologically-conscribed part of the lexicon, and are subject to an assimilatory or featural-changing process that acts on natural classes of segments. In other words, they look like the work of a phonological grammar, and capturing them with extra-grammatical means only to maintain a gradual progression view seems somewhat arbitrary. A second crucial point is that these three regressions are 'pure' in the sense of Stemberger et al (1999): in moving from one stage to the next, one aspect of the child's output gets *further* from the target, and at the same time no other property gets *closer*. With these properties in mind, here is a definition of the phenomenon:

(3) <u>Grammatical backtracking (definition)</u>
A change in rankings from stage x to stage y, whereby the output for an input /I/ at stage y incurs a proper superset of the faithfulness violations incurred by /I/'s output at stage x

To connect this definition with the data we have seen, compare the potential for nasal harmony in the word *down* at stages one and five of Daniel's development from (1). At stage one, *down* has a faithful initial consonant so its winner at this stage x incurs no violations of a faith constraint like IDENT[NASAL], as in (4a) below. But by stage five the [d] of *down* has succumbed to harmony with the final [n] to obey a constraint like AGREE[NASAL] (see section 2.4), meaning that IDENT[NASAL] is now violated (4b). And

[1] In this notation: 'K' refers to all velar obstruents; 'T' to all coronals, 'P' to all labials and 'V' to all vowels.

since this winner at stage y does not satisfy any faith constraints better than at stage x, the progression from (4a) to (4b) counts as a grammatical backtrack.

(4) Grammatical backtracking in nasal harmony from (1), as defined in (3)

(a) Daniel at stage one

/daʊn/	IDENT[NAS]	(MARKEDNESS)
☞daʊn		(*)
naʊn	*!	

(b) Daniel at stage two: backtracking

/daʊn/	AGREE[NAS]	IDENT[NAS]	(MARKEDNESS)
daʊn	*!		(*)
☞naʊn		*	

Since this type of grammatical backtracking is indeed attested in children's development – is the gradual progression of phonological development necessarily wrong? Is there anything common to the attested cases of grammatical backtracking that suggest a solution? Section 2.3 argues that there is.

2.3 *The source of grammatical backtracking* The illustration of backtracking between (4a) and (4b) involves the appearance of a constraint AGREE[NASAL] at the latter stage. Where did this constraint come from? Was it somehow promoted higher in the learner's ranking compared to previous grammars? On what grounds?

An alternative view, which is certainly not new to the present work, is that this AGREE[NASAL] appears at the top of the hierarchy in (4b) because it is novel to the grammar – that is, it has been induced by the learner, later than the tableau in (4a). This is the suggestion in Becker and Tessier (2011) with respect to Trevor's place harmony, where it is noted that this process is widely attested in child phonologies, but not in target adult phonologies (e.g. Vihman, 1978). Similarly, positional velar fronting (as in the regression of (2)) is also a child-specific process, not observed in adult languages; velar fronting is also reported to regress across the lexicon in other children's development, and has been argued to derive from child-specific articulatory pressures (Inkelas and Rose, 2008; McAllister Byun, 2011).

Thus, the strongest claim is that grammatical backtracking is always the result of a child-specific constraint, induced after the earliest stages of production – predicting that children's obedience to markedness constraints which also constrain adult phonologies should *not* show grammatical backtracking, under the assumption that they are present in all typically developing phonologies from the beginning.[2] Of course, much more careful study of longitudinal data will be necessary to strengthen or reject this empirical claim. It will also be necessary to propose overt mechanisms by which a learner could come to induce a constraint at some time *after* the onset of output phonology; see also the related proposals of McAllister Byun (2011) and Inkelas, McAllister Byun and Rose (2012).

In any event: grammatical backtracking is not the only kind of U-shape seen in phonological development. Below are a couple examples of regression that do *not* fall into the category of grammatical backtracking, and why.

2.4 *Sources and profiles of U-shapes that are not grammatical backtracking* Looking at the full range of children's phonological development, there are many other trajectories that show regressions in one or more senses but do not fit the definition of grammatical backtracking. Some of these may come about through the normal application of the learner's error resolution mechanisms, and others will require explanations that fall outside the learner's production grammar.

Representative examples come from data in Stemberger, Bernhardt and Johnson (1999), tracking one child Morgan's progress and regress while learning segmental contrasts in word-final position. At an initial stage between 1;4-1;6 Morgan unrounded all word-final vowels; then at 1;6.22, she moved into a stage where these unrounded vowels were produced followed by an additional [m]:

(5) Regression at the ends of words (data from Stemberger *et al*, 1999)

	1;4-1;6	1;6.22
no	nɤː	nɤː ~ nɤːm
shoe	ɕɯː	ɕɯː ~ ɕɯːm
cow	tʰaɯ	tʰaɯ ~ tʰaɯm

[2] Whether innate or induced from experience in some language-independent way – a topic far too wide to tackle here.

Both of these stages can be understood as the influence of a grammar which does not tolerate rounded vowels (for the sake of descriptive simplicity the analysis will treat all round vowels, although Morgan is reported to have imposed this pattern only the ends of words.) At the first stage, her grammar avoids vowel rounding by deleting the vowel's rounding feature, violating MAX[LABIAL]. At the second stage, this labial feature is now faithfully retained, but on an adjacent epenthetic segment; this process of featural 'fission' violates the faithfulness constraint UNIFORMITY (McCarthy and Prince, 1995).[3]

Moving from this first stage to the second can be achieved using a variety of different error-driven learners within Optimality Theory; as a certain type of learner will be crucial to the discussion in section 4, the process will now be sketched here. At the first stage, the learner will attempt to produce forms like 'no' by feeding them as *input* to their grammar as in (6a), and thereby make the error in (6b), represented in the ERC format of Prince (2002) and others. This ERC tracks whether each constraint prefers the target 'winner' form in (6a), being fully-faithful to the (observed) input, or whether it prefers the current grammar's output in (6a), which in the eventual target grammar will be a 'loser':

(6)

(a) Stage one: loss of vowel rounding at 1;4[4]

/no$_1$/	*ROUNDV	UNIF	MAX[LAB]
no	*!		
☞nɤ:			*
nɤ$_1$:m$_1$		*!	

(b) Analyzing the error at 1;4

W VS L	*ROUNDV	UNIF	MAX[LAB]
[no] vs. [nɤ:]	L	e	W

Result of (6b): *MAX[LAB] >> *ROUNDV

The kind of error-driven learner used here (either the BCD algorithm of Prince and Tesar, 2004, or the LFCD algorithm of Hayes, 2004) builds new rankings with two goals: first to resolve the errors in their data, like (6b), and second to keep their grammar as restrictive as possible. To meet the first goal, the learner installs a winner-preferring constraint above all loser-preferring constraints (Prince and Smolensky, 2004) – in this case, this uniquely chooses the new ranking MAX[LAB] >> *ROUNDV. To meet the second goal the learner imposes ranking biases, starting with the drive to rank {all M} >> {all F} unless errors prove otherwise – in this case, *ROUNDV >> UNIFORMITY, since the error in (6) gives the learner no data on the role of UNIFORMITY in the target grammar. With these two rankings in place, the learner's new grammar is (7):

(7) Stage two: labial fission at 1;6 [**regression**]

/no$_1$/	MAX[LAB]	*ROUNDV	UNIF
no		*!	
nɤ:	*!		
☞nɤ$_1$:m$_1$			*

(b) Analyzing both errors:

W VS L	*ROUNDV	UNIF	MAX[LAB]
no vs. nɤ:	L	e	W
no vs nɤ$_1$:m$_1$	L	W	e

This regression is thus a trade-off between methods of avoiding round vowels, created by the normal workings of this learning algorithm; the new grammar in (7) improves faithfulness to the rounding but decreases faithfulness to input segmental count. Note that the next learning cycle, when both errors in (7b) are resolved, will bring the learner to the target grammar which faithfully preserves round vowels.

On the other hand, Morgan's treatment of round vowels at stage two also included some outputs that are not explained by the grammar in (7a). As shown in (8a), a few of Morgan's words with final unrounded /i/ were also produced (sometimes variably) with a final [m]; conversely, at least one word with a final round vowel was not produced with final [m] (8b):

3 An example of a similar featural fission process in adult phonology is reported in Crowley (2004) for French borrowings into Bislama, a creole of Vanuatu, in which French nasal vowels inputs are produced as oral nowels + nasal consonant: /ka'mjõ/ → [kamioŋ].

4 For the sake of keeping the input and outputs legible among the candidate indices, the target vowel has been transcribed as merely [o] instead of the rather more correct diphthong [oʊ] or similar.

(8) Lexically-specific exceptions among word-final vowels

		1;4-1;6	1;6.22
a)	*me*	mi:	mi: ~ mi:m
	sockie	laki	lakim
	doggie	dagi	dagim

		1;4-1;6	1;6.22
b)	*yellow*	lalɤ:	lalɤ:, * lalɤ:m

The [m]-insertion pattern in (8a) is clearly a regression, and while it does not look like any attested grammatical patterns in adult phonologies, it also does not appear to have any phonological basis: there is no ready explanation for why the items in (5) and also (8a) would receive epenthetic final [m]s, but not other word-final vowels, including (8b).

The conclusion drawn here is that regressions like (8a) are not grammatically-conditioned. Rather, it is suggested that they represent a kind of lexical 'leakage', in which the grammar's mapping in (7) to add [m] to the outputs of certain words has overwritten onto the inputs of a few other lexical items. This is not a novel suggestion: see for example the related proposals discussed in Menn and Matthei, (1992) and references therein. It should also be noted that regressions may come from a variety of places outside the production grammar – for instance, see Macken (1980)'s arguments that regressions can be caused by the resolution of perceptual errors, causing overcorrection of existing inputs.

Leaving aside the source of other regression types (as unfinished as the discussion is), the rest of this paper returns to the notion of 'trade-offs' between constraints, raised in the prose below (7). The focus will be a kind of trade-off that does not appear to occur, and its learning consequences.

3 An unattested phonological, grammatical regression: process trade-off

3.1 *Defining process trade-offs* The previous section illustrated an example of an observed regression (exs. 5-7) which could come about in the grammar via normal error-driven learning; as schematized in (9a) below, it came about from a reversal in the ranking of two faithfulness constraints. A different kind of grammatical regression is sketched in (9b), in which two M vs. F rankings are reversed – so that at stage one, only M1 is obeyed (causing some process violating F1), while at stage two M1 is now violated but M2 is obeyed (through violation of F2) . This kind of reversal in (9b) will be referred to as a *process trade-off.* [5]

(9a) Attested regression via learning:
Stage one: M >> **F1** >> **F2**
Stage two: **F2** >> M >> **F1**

(9b) Unattested regression: a process trade-off
Stage one: **F2** >> M2, M1 >> **F1**
Stage two: **F1** >> M2, M1 >> **F2**

From a general cognitive perspective, it might make sense to expect process trade-offs to occur in the course of phonological development. If children tackle one difficult phonological structure at a time, they might sometimes regress with respect to something otherwise-acquired while shifting their focus to something else hard and new; we will see in the next section how this characterization matches the re-ranking illustrated in (9b). However, the present claim is that, unlike (9a) and other type of regressions seen in section 2, process trade-offs like (9b) are unattested in children's typical phonological acquisition. If this empirical observation holds up, then it requires that any learning algorithm aimed at capturing children's phonological acquisition avoid this grammatical regression (to which we return in sections 4-5).

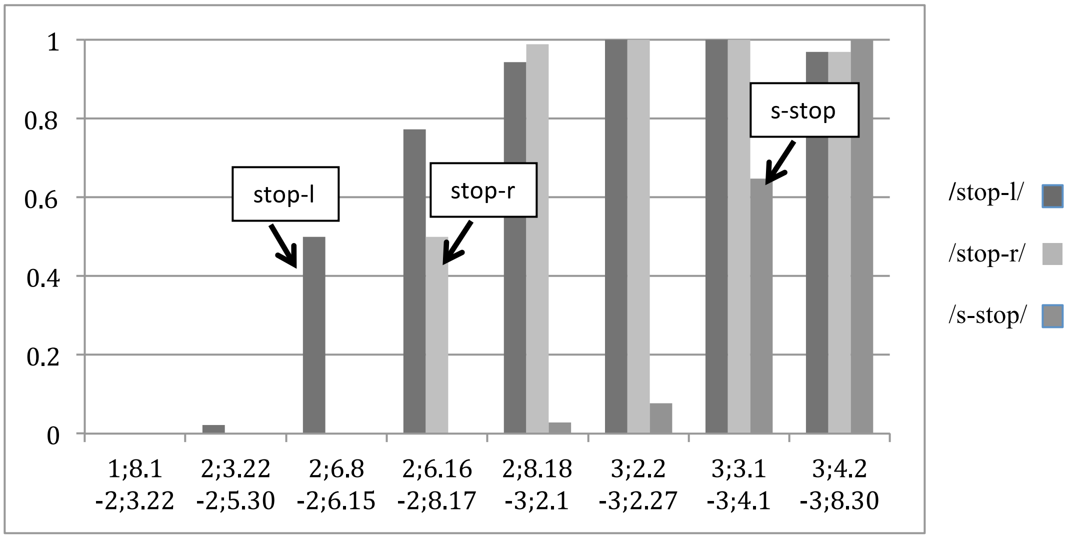
3.2 *One search for process trade-offs* As one attempt to search for potential process trade-offs (spoiler alert: the search failed), a case study was conducted analyzing a diary study corpus of one child Zack, provided in an appendix to Smith (2010).[6] From this corpus, all transcribed tokens containing target two consonant onset clusters were extracted; on the basis of these data, it was decided to focus on three

[5] Process tradeoffs as defined here are distinct from and orthogonal to 'cumulativity' effects in child speech, discussed in Farris Trimble (2008), Jesney (2011), Levelt and van der Veijer (2004) and elsewhere, in which a child appears to tolerate multiple marked structures on their own, but not all within a single output, particular at a single locus. Note for example that these cumulative effects hold at a single stage of a child's grammar, whereas trade-off represent a series of developmental stages.

[6] Special thanks to Philip Dilts for transcription and data manipulation.

types of onset cluster: /stop+r/, /stop+l/ and /s+stop/. A hand-crafted analysis determined a set of 8 stages between ages 1;8 and 3;4, during which each cluster type moved from nearly all reduction (/CC/ → [C]) to cluster preservation (/CC/ → [CC]).[7] As shown in (10) below, /stop-l/ clusters began to surface faithfully during the week of 2;6.8-15, while /stop+r/ clusters began their gradual appearance a week later (between 2;6.15 and 2;8.17) and /s+stop/ clusters finally began to be produced in earnest during the month 3;3 to 3;4.

(10) Proportion of /CC/ onset tokens produced as clusters at each stage (data from Smith, 2010)



Having determined these three stages, all the extracted words were examined which occurred in the corpus both before and after the emergence of each cluster, looking for other marked structures that could have 'flipped'. For example, in the first three stages of (10), Zack produced 72% of his input coda fricatives as stops (31/43 up until 2;6.15), but then they began to be produced much more often as fricatives (only 43% stopped, 12/28, during stage 4). Thus the acquisition of stop-r and stop-l clusters, combined with the acquisition of coda fricatives, could have provided an opportunity for a process trade-off, as in (11):

(11) <u>Hypothetical process trade-off</u>

(a) Stage one: Coda fricatives acquired

/pliz/	**IDENT [STOP]**	*STOP-L ONSET[8]	*CODA FRIC	**MAX**
pid	*!			*
☞piz			*	*
plid	*!	*		
pliz		*!	*	

(b) Stage two: Clusters acquired, codas **regress**

/pliz/	**MAX**	*STOP-L ONSET	*CODA FRIC	**IDENT [STOP]**
pid	*!			*
piz	*!		*	
☞plid		*		*
pliz		*	*!	

Nevertheless, in the corpus of 2620 tokens with target biconsonantal onsets, no such process trade-offs were found. A sample of words transcribed at 2;6 and 2;9 in (12a) provides the kind of progress seen instead: one structure appears and then the other, and while both relative orders are observed, no process tradeoffs are seen (indicated as ungrammatical forms with *).

The items in (12b) deal with another potential source of trade-offs after the emergence of /s-stop/ onsets during the month of 3;3. From the beginning of the corpus up until 3;4.1, virtually every velar stop was fronted to a coronal (e.g. 34/36 tokens were still fronted during the month of 3;3). Starting at 3;4.2 this fronting became much less common, with only 41% of velars fronted – 50 out of 123 tokens – between 3;42 and 3;8.30. Again, however, (12b) shows how the acquisition of /s+stop/ clusters and velar stops did

[7] All that is crucial in this faithful mapping is that *some* cluster appear in outputs; whether the cluster is segmentally accurate is a different matter.
[8] This constraint should be understood as part of a family of constraints against particular sonority profiles in onset – as in e.g. Baertsch (2002) or others.

not cause any process trade-offs in any words seen before and after these transitions:

(12) No process trade-offs during Zack's onset cluster acquisition

(a) **stop-l onsets** vs. coda frics			
	@ 2;6	@ 2;9	*unattested trade-off*
Gruff	[dʌf]	[dɹʌf]	*[dɹʌp]
close	[trəud]	[trəuz]	*[təuz]
please	[pid], [piz]	[pɹiz]	*[prid]

(b) **s-stop onsets** vs. velars			
	@ 3;3	@ 3;5	*unattested trade-off*
scoop	[stu:p]	[sku:p]	*[ku:p]
sky	[stai]	[stai], [skai]	*[kai]
stick	[stɪt]	[stɪt], [stɪk]	*[tɪk]

The upshot of this study of Smith (2010)'s corpus is simply that no process trade-offs were attested in Zack's acquisition of onset clusters, compared with any other marked structure. If this result turns out to be representative, then the kind of learner already used in section 2.4 will need some improvement – because as the next section shows, it is indeed prone to such regressions.

4 The danger of process trade-offs in error-driven learning

4.1 *Walking through the creation of a process trade-off* There are two key properties of an OT error-driven learner that contribute to an unwanted creation of regressions like process tradeoffs. The first is cumulative learning: that is, storing a set of reliable errors from which to learn as in (7b), and each time building a ranking from scratch. The second is a persistent ranking bias which ensures M >> F rankings in part by trying to install F constraints that 'free up' as many M constraints as possible (explained below). The following example shows how an OT learner with these properties can regress when exposed just to a single learning datum like *please,* using the constraints already used in (11). Note that the details of how the learner's algorithm gets from ERCs to rankings will be highly abbreviated here, so the reader is pointed to Prince and Tesar (2004) and Hayes (2004) for more detail.

With a learner biased to rank all M >> all F until data prove otherwise, the initial state of our learner will be maximally restrictive: using the ranking in (13a) to create the ERC in (13b):

(13a)

/pliz/	*Stop-L onset	*Coda Fric	Max	Ident [Stop]
☞ pid			*	*
piz		*!	*	
plid	*!			*
pliz	*!	*		

(13b)

W vs L	*Stop-L onset	*Coda Fric	Max	Ident [Stop]
pliz ~ pid	L	L	W	W

Recall that the learner's first goal is to resolve its set of errors, and that this means installing *some* W-preferring constraint above *all* L-preferring constraints. In (13b), this means that *either* faith constraint will end up ranked highest; with only this piece of data there is no way to choose between them, so suppose the learner picks Ident. Installing Ident at the top of the ranking will resolve (13b) – because no matter what else happens, this new grammar already prefers faithfully mapping /z/ → [z], not *[d] – so the ranking biases can now freely impose the learner's second goal of keeping {M} >> {F}. As a result both Markedness constraints get installed next, and Max at the bottom.

This new ranking, given in (14), is also Stage one from (11a); as shown there, it will map /pliz/ to the slightly-improved output [piz], and add a new error to the ERC set, now added in (15):

(14) Ranking learned from (13b): Ident[Stop] >> *Stop-L onset, *CodaFric >> Max

(15) ERCs at second round of learning (combining 13b and 11b):

input	W vs L	*Stop-L onset	*CodaFric	Max	Ident [Stop]
/pliz/	pliz ~ pid	L	L	W	W
	pliz ~ piz	L	e	e	W

To learn from the sum of (15), the learner again begins by installing some W-preferring constraint above all the L-preferring constraints – but this time, the two errors *do* provide an asymmetry between the two faithfulness constraints that assign Ws. Keeping in mind the second goal of maximizing {M} >> {F}, the learner's biases now prefer to install MAX at the top of the hierarchy. This will resolve both errors in one fell swoop, freeing up the learner to install both markedness constraints in the next stratum, and ending up with the ranking in (16). However – this new grammar is now Stage two from (11b), which created a process trade-off! This learner's three stages of learning including the regression are summarized in (17):

(16) Ranking learned from (15): MAX >> *STOP-L ONSET, *CODAFRIC >> IDENT[STOP]

(17) Three stages leading to regression:

1. /pliz/ → [pid] (grammar in 13a)
2. /pliz/ → [piz] (grammar in 11a)
3. /pliz/ → [plid] (grammar in 11b) – *when clusters are learned, codas regress*

4.2 *Diagnosing the root cause of process trade-offs* What should be learned from the preceding section's result? That is, what kind of solution should we seek to avoid unattested regressions? A few different options are possible. One type of solution is to abandon the ranking-from-scratch and the persistent ranking biases of BCD or LFCD, in favour of a gradual algorithm like the GLA (Boersma, 1998) which makes small changes to the grammar at each learning cycle but does not retain any of its errors. A GLA-like learner will never move in one learning cycle between rankings as different as stages 2 vs. 3 in (17), and it never assesses two errors simultaneously as in (15).

An alternative solution, however, is driven by a diagnosis of exactly what causes these learners regress -- and the ultimate source is the indeterminacy of the very first stage's error in (13a). The resulting ERC in (13b) has two Ls and two Ws, but the learner's algorithm has no way of knowing which Ws and Ls are connected, so it must pick a ranking, make another error and stumble on. This indeterminacy is directly related to Dresher (1999)'s notion of the 'credit problem' when learning from errors. In this circumstance, adding the (13a) ERC to the learning data can easily force the learner to try giving credit to different constraints each in turn across successive stages, and thus causing the flip-flopping regression of a process trade-off.

Given this analysis, the final section below considers a solution which seeks to avoid unattested regressions by not storing errors like (13), using a non-OT grammar.

5 Avoiding process trade-offs in Harmonic Serialist Learning

5.1 *A crash course in Harmonic Serialism* This section cannot hope to do robust justice to the entire phonological framework of Harmonic Serialism – originally suggested in Prince and Smolensky (2004: 94-95), but proposed in earnest in McCarthy (2007) and many subsequent works. The reader who is unfamiliar with HS will only get the bare minimum with which to grasp the idea sketched here, but see McCarthy (2008ab) and other references in this section for real introductions and applications.

Compared to OT, there are two crucial distinguishing features of HS. The first is that it maps from input to output via a derivation, i.e. a series of gradual steps in which each output is taken as input to the next step. Like OT, every language uses the same ranking of constraints in every mapping; but the second difference is that inputs change gradually in every mapping because the candidate set is very limited. In particular, the candidate set consists of the fully-faithful candidate, and candidates which are each 'one step away' from the input; for current purposes, we can define 'one step away' as meaning a candidate which incurs only <u>one</u> violation of a core faithfulness constraint: MAX, DEP, IDENT[F] and so on.[9] Thus, the first HS tableau for *please* in our running example above looks like (18) below:

[9] This definition is limited to 'core' faithfulness constraints so to not include e.g. positional faithfulness constraints. Since any violation of IDENT[ONSET] is also a violation of general IDENT, the theory must ensure that both violations are not counted in building candidates, or else unfaithfulness to an onset could never be achieved.

(18) HS initial state derivation: step one

/pliz/	*STOP-L ONSET	*CODA FRIC	MAX	IDENT [STOP]	
~~pid~~					*not in the candidate set!*
☞piz		*	*		*violation of Max*
plid	*!			*	*violation of Ident[stop]*
pliz	*!	*			*fully-faithful candidate*

While this tableau has an initial state, {all M} >> {all F} ranking just as in (13), the HS grammar cannot choose the maximally-unmarked [pid] here: since /pliz/ → [pid] would require *two* faithfulness violations, [pid] is not even in /pliz/'s candidate set. Instead, the winner takes one step, satisfying the onset cluster constraint and surfacing as [piz]. In (19), this form is now taken as the input for a second step – same ranking, new candidate set, new winner:

(19) HS initial state derivation: step two

/piz/	*STOP-L ONSET	*CODA FRIC	MAX	IDENT [STOP]	
☞pid				*	*violation of Ident[stop]*
piz		*!			*fully-faithful candidate*

Now that /pid/ is only one step away from the input, it is included in the candidate set, and so it wins by satisfying *CODAFRICATIVE. And when this winner is again fed to the same ranking for a third round, it maps to the fully-faithful candidate – and so the derivation is complete:

(20) HS initial state derivation: step three, and convergence

/pid/	*STOP-L ONSET	*CODA FRIC	MAX	IDENT [STOP]	
☞ pid					*fully-faithful candidate*
piz		*!		*	*violation of Ident[stop]*

In this way, this grammar maps /pliz/ → piz → [pid], using an intermediate step to get from input to output.

HS has been argued to solve several different drawbacks of fully-parallel OT, including ways to capture opaque interactions, avoid too-many-repairs problems, and others (see e.g. Jesney, 2008; Kimper 2011; McCarthy 2008a,b; Pruitt 2010.) In the next section, I suggest a way in which learning via HS can also avoid process trade-offs.

5.2 *How the HS learner avoids process trade-offs* An HS approach to learning can indeed be very similar to the one used within OT in section 4, with one small adjustment which turns out to have important consequences. The crucial difference is in deciding what to store in an ERC: how to compare the winners and losers.

The ERCs used for OT learning as in (15) were all comparisons between the observed form being learned, taken as the input, and the current grammar's output; Prince (2002) and others provide the formalization of how to map from a tableau to an ERC's vector of Ws, Ls and es, But notice that in the derivation for *please* in HS, the input /pliz/ and the output [pid] do not co-occur within a single tableau: /pliz/ is found in (18), while /pid/ and [pid] are in (19)-(20). Notice too that in the tableaus where the intended loser [pid] does occur, the faithfulness violations incurred are with respect to an input *other than the intended winner*. Thus, the HS tableaus for a multi-step derivation do not provide straightforward data with which to build an ERC comparing the input and ultimate output.

Various options might be taken for constructing HS learning ERCs – but the one adopted here, also used in Jesney and Tessier (to appear), is to take the *first derivation* of any mapping to build an ERC. In this case, that means just using the tableau in (18), mapping /pliz/ initially to [piz], to create the ERC in (21). Note that because it was built from only one step of an HS derivation, this ERC has *only one W-preferring constraint*: the result of the one faithfulness violation that the grammar chose as optimal at step one.

(21) ERC from first step of HS derivation in (18)

input	W vs L	*STOP-L ONSET	*CODAFRIC	MAX	IDENT [STOP]
/pliz/	pliz ~ piz	L	e	W	e

From this ERC, the learner will resolve their error by ranking the W-preferring MAX constraint above *STOP-LONSET, moving from the initial state to a second stage as in (22) below:

(22) Ranking learned from (21): MAX >> *STOPLONSET, *CODAFRIC >> IDENT[STOP]

This grammar, shown at work in (23) below, is another stage x in the process trade-off we saw in section 3: one M constraint is no longer obeyed, so the output is slightly more target like (clusters are retained) but the other M constraint is still causing errors (stopping coda fricatives):

(23) Derivation one, using ranking from (22)

/pliz/	MAX	*STOP-L ONSET	*CODA FRIC	IDENT [STOP]	
~~pid~~					*not in the candidate set!*
piz	*!		*		*violation of Max*
☞plid		*		*	*violation of Ident[stop]*
pliz		*	*!		*fully-faithful candidate*

When run through a second step, this winner /plid/ will be mapped to itself, and so this grammar converges on [plid] as its final output:

24) Derivation two and convergence

/plid/	MAX	*STOP-L ONSET	*CODA FRIC	IDENT [STOP]	
pid	*!		*		*violation of Max*
☞ plid		*			*fully-faithful candidate*
pliz		*	*!	*	*violation of Ident[stop]*

Again the learner will build an ERC from the first step of the current grammar's mapping, i.e. (23), and add it to the previously stored error, as in (25):

(25) ERCs from first steps of HS derivations in (18) and (23)

input	W vs L	*STOP-L ONSET	*CODAFRIC	MAX	IDENT [STOP]
/pliz/	pliz ~ piz	L	e	W	e
/pliz/	pliz ~ plid	e	L	e	W

Resolving these two errors requires a grammar with two separate F >> M rankings – and once these are established (in either relative orders), the learner will have reached the target grammar as in (26):

(26) Ranking learned from (25) MAX >> *STOPLONSET >> IDENT[STOP] >> *CODAFRIC OR
IDENT[STOP] >> *CODAFRIC >> MAX >> *STOPONSET[10]

Thus, the HS learner has gone through three stages from initial to target, with no process trade-offs:

[10] A complicating factor is being ignored here: that *CODAFRIC >> MAX in fact predicts that coda fricatives are deleted rather than stopped, i.e. /pliz/ → [pli]. If this ranking is learned from (25), then, a further learning cycle will be necessary to re-rank MAX >> *CODAFRIC before a grammar consistent with the target is reached.

(27) Three stages leading to target, without regression!

1. /pliz/ → [pid] (grammar in 18-20)
2. /pliz/ → [plid] (grammar in 23-24)
3. /pliz/ → [pliz] (grammar in 26) – *target!*

To re-iterate: the fact that this learner did not regress comes from the simplified nature of its ERCs. Because they are always taken from a single HS tableau, the difference between their winner and loser will always be a single core faithfulness violation; they will therefore not contain the multiple Ws of the OT ERC in (13b) and (15) that caused the learner to flip-flop between rankings. By learning a single re-ranking from each error, the learner avoids the kind of wholesale grammar re-organization that can cause unexpected regressions.

6 Conclusions

This paper has drawn together several strands of research on phonological acquisition and learnability. The goals have been (i) to draw a connection between child-specific phonological processes and the potential for true U-shaped grammatical development, as distinguished from other kinds of regressions; (ii) to connect this U-shape with a learner's ability to induce new constraints later in learning; (iii) to characterize a type of unattested grammatical regression, and demonstrate its unattestedness in a corpus; and (iv) to assess the likelihood of such unattested regressions in two error-driven, biased learning approaches, concluding that a parallel OT learner can easily regress, while the serial and gradual Harmonic Serialist learner does not. Taken altogether, this paper's larger goal is to encourage more analyses of children's U-shaped phonology, in any framework, and to argue that comparing messy child data and formal learning approaches can be fruitful for our understanding of both.

References

Baerstch, Karen. 2002. *An Optimality-Theoretic approach to syllable structure: The Split Margin Hierarchy*. Ph.D. Dissertation, Indiana University.

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

Bernhardt, Barbara and Joseph Stemberger 1998. *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

Bleile, Ken & Bruce Tomblin 1991. Regressions in the Phonological Development of Two Children. *Journal of Psychological Research* 20(6): 483-99.

Boersma, Paul. 1998**.** *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam.

Compton, A.J and M. Streeter. 1977. Child phonology: data collection and preliminary analyses*. Papers and Reports on Child Language Development* 13: 99–109.

Chiat, Shulamuth. 1983. Why *Mikey's* right and my *key's* wrong: The significance of stress and word boundaries in a child's output system. *Cognition* 14: 275-300.

Crowley, Terry. 2004. *Bislama Reference Grammar.* Oceanic Linguistics Special Publications, No. 31. University of Hawai'i Press.

Dresher, B. Elan. 1999. Charting the Learning Path: Cues to parameter settings. *Linguistic Inquiry* 30(1): 27-67.

Farris-Trimble, Ashley. 2008. *Cumulative faithfulness effects in phonology.* Ph.D. dissertation, Indiana University.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. In Kager et al. (2004). 158–203.

Inkelas, Sharon, McAllister Byun, Tara and Yvan Rose. 2012. Transient constraints and phonological development. Poster presented at the International Child Phonology Conference, Minneapolis, MN.

Inkelas, Sharon and Yvan Rose. 2008. Positional neutralization: a case study from child language. *Language* 83: 707-736.

Jesney, Karen. 2008. Positional Faithfulness, Non-Locality and the Harmonic Serialism Solution. In S. Lima, K. Mullins and B. Smith (eds). *Proceedings of NELS39.* Amherst, MA: GLSA publications.

Jesney, Karen. 2011. *Cumulative constraint interaction in phonological acquisition and typology.* Ph.D. dissertation, UMass Amherst.

Jesney, Karen and Anne-Michelle Tessier, to appear. Learning in Harmonic Serialism and the necessity of a richer base. To appear in *Phonology*.

Kager, René, Joe Pater & Wim Zonneveld (eds.) 2004. *Constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Kimper, Wendell. 2011. Locality and Globality in Phonological Variation. *Natural Language and Linguistic Theory 29(2)*: 423-465.

Levelt, Claartje and Ruben van der Vijver. 2004. Syllable types in cross-linguistic and developmental grammars. In Kager et al (2004).

Macken, Marlys A. 1980. The child's lexical representation: the 'puzzle-puddle-pickle' evidence. *Journal of Linguistics* 16: 1-17.

McAllister Byun, Tara. 2011. A gestural account of a child-specific neutralisation in strong position. *Phonology, 28:* 371-412.

McCarthy, John J. 2007. Restraint of Analysis. In E. Bakovic, J. Ito and J. J. McCarthy (eds.) *Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince.* Linguistics Research Center, UC Santa Cruz. Available at: http://escholarship.org/uc/item/8j31x9md.

McCarthy, John J. 2008a. The gradual path to cluster simplification. *Phonology* 25(2): 271-319.

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

McCarthy, John J. and Alan Prince. 1995. Faithfulness and reduplicative identity**.** In J. Beckman, S. Urbanczyk and L. Walsh Dickey (eds.) *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory:* 289-384. Amherst, MA: GLSA Publications.

Menn, Lise. 1971. Phonotactic rules in beginning speech. *Lingua* 26: 225-251.

Menn, Lise. 1983. Development of articulatory, phonetic, and phonological capacities. In B. Butterworth (ed). *Language Production Vol 2.* London: Academic Press.

Menn, Lise and Edward Matthei. 1992. The 'two-lexicon' account of child phonology: looking back, looking ahead. In Charles A. Ferguson, Lise Menn & Carol Stoel-Gammon (eds.) *Phonological development: models, research, implications.* Timonium, Md.: York Press. 211–247..

Pater, Joe 1997. Minimal violation and phonological development. *Language Acquisition* 6: 201–253.

Pater, Joe and Jessica A. Barlow 2003. Constraint conflict in cluster reduction. *Journal of Child Language* 30: 487–526.

Prince, Alan. 2002. Entailed Ranking Arguments. Ms., Rutgers University. New Brunswick, NJ. [ROA-500].

Prince, Alan and Bruce Tesar. 2004. Learning phonotactic distributions. In Kager et al. (2004). 245–291.

Pruitt, Kathryn. 2010. Serialism and locality in constraint-based metrical parsing. *Phonology* 27(3), 481-526.

Smith, Neilson.V. 2010. *Acquiring phonology: A cross-generational case study.* Cambridge: CUP.

Staubs, Robert and Joe Pater (to appear). Learning serial constraint-based grammars. To appear in J. J. McCarthy and J. Pater, (eds.) *Harmonic Grammar and Harmonic Serialism.* London: Equinox Press.

Stemberger, Joseph P., Barbara H. Bernhardt, and Carolyn E. Johnson. 1999. "Regressions" ("U"-shaped learning) in the acquisition of prosodic structure. Poster presented at the 6th International Child Language Congress. Handout available at: *roa.rutgers.edu/files/471-1101/471-1101-STEMBERGER-0-0.PDF*

Tessier, Anne-Michelle. 2012. Error-driven learning in Harmonic Serialism. To appear in the proceedings of NELS 42: 545-558.

Vihman, Marilyn 1978**.** Consonant harmony: its scope and function in child language. In J. H. Greenberg, C. A. Ferguson & E A. Moravcsik (eds.) *Universals of human language. Vol. 2: Phonology.* Stanford: Stanford University Press. 281–334.

Vihman, Marilyn and Shelley Velleman. 1989. Phonological re-organization: a case study. *Language and Speech* 32(2): 149-170.

Vihman, Marilyn and Shelley Velleman. 2002. Whole-word Phonology and Templates: Trap, Bootstrap or Some of Each? *Language, Speech and Hearing Services in the Schools* 33: 9-23.