Learning prevents MaxEnt from giving probability to harmonically bounded candidates

Charlie O'Hara
University of Southern California

AMP 2019
October 13, 2019

Introduction

A major goal of phonological theory is to develop a model that can capture the attested phonological patterns while not vastly over-predicting.

- Constraint based grammars (Optimality Theory[1], Harmonic Grammar[2], etc.) make strong typological predictions through **Factorial Typology**
- Recently, an abundance of work[3] has investigated the hypothesis that learnability affects both categorical and soft typology.

[1] Prince & Smolensky (1993/2004); McCarthy & Prince (1995)

[2] Legendre *et al.* (1990); Pater (2016)

[3] Boersma (2003); Pater & Moreton (2012); Staubs (2014); Hughto (2018); O'Hara (2017, in prep, 2018, 2019)

Introduction

A major goal of phonological theory is to develop a model that can capture the attested phonological patterns while not vastly over-predicting.

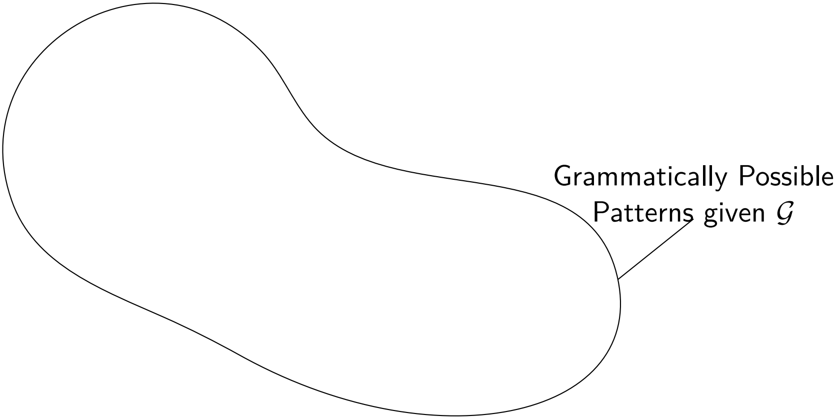
- Constraint based grammars (Optimality Theory[1], Harmonic Grammar[2], etc.) make strong typological predictions through **Factorial Typology**
- Recently, an abundance of work[3] has investigated the hypothesis that learnability affects both categorical and soft typology.

[1] Prince & Smolensky (1993/2004); McCarthy & Prince (1995)

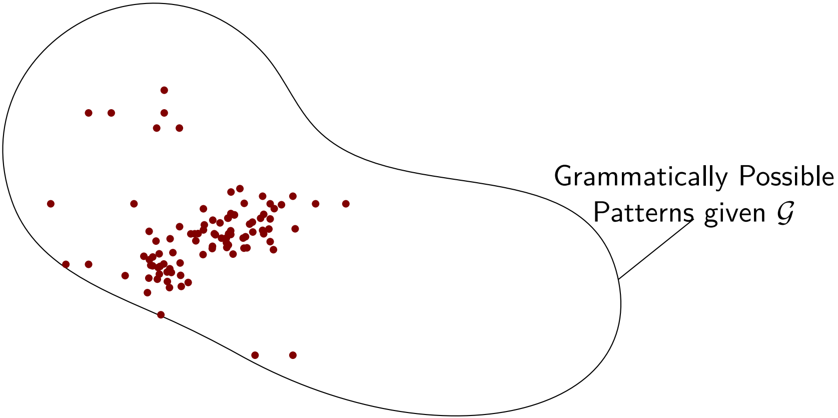
[2] Legendre *et al.* (1990); Pater (2016)

[3] Boersma (2003); Pater & Moreton (2012); Staubs (2014); Hughto (2018); O'Hara (2017, in prep, 2018, 2019)

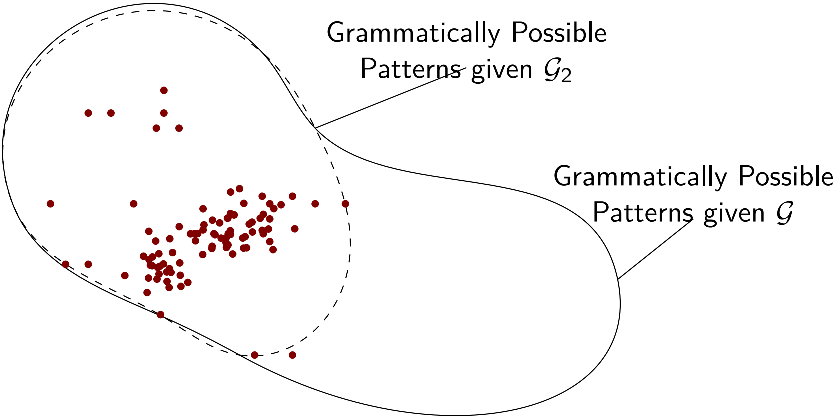
Restrictiveness and Learning



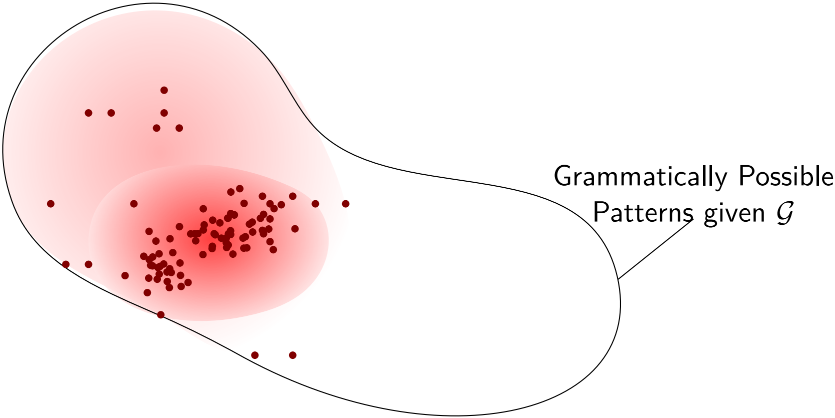
Restrictiveness and Learning



Restrictiveness and Learning



Restrictiveness and Learning



Simply Harmonically Bounded Candidates

A candidate is SIMPLY HARMONICALLY BOUNDED by another candidate if it has a proper superset of the violations of that candidate.

- /CV/→[V] is simply harmonically bounded by /CV/→[CV]

/CV/	DEP	MAX	ONSET
☞ a. CV			
b. V		-1	-1

- In Classic OT, Categorical HG, and Noisy HG, a harmonically bounded candidate will never surface.
- With these constraints, no ranking/weighting is able to find a pattern where onsets delete.

MaxEnt and Harmonically Bounded Candidates

But in MaxEnt, simply harmonically bounded candidates can receive probability (Jesney, 2007).

- As a result, MaxEnt can over-generate categorical (and noisy) HG (see also Anttila & Magri (2018)).
- As an example, MaxEnt generates a pattern where onsets variably delete.

/pa/ → [pa] 50%, [a] 50%   /u/ → [u] 100%

	$w = 10$	$w = 0$	$w = 0$		
/CV/	Dep	Onset	Max	Harm	Prob
a. CV				0	.5
b. V		-1	-1	0	.5
/V/	Dep	Onset	Max	Harm	Prob
c. CV	-1			-10	~ 0
d. V		-1	-1	0	~ 1

Problem

MaxEnt can model grammars with probabilistic markedness reversals.

- It is impossible to restrict typology in MaxEnt using just the grammar.
- When all weights equal zero, all candidates for each input receive the same probability.
- Any input output mapping can receive probabilty

Not ALL grammars are equally likely.

- Patterns that give substantial probability to harmonically bounded candidates are much harder to learn than patterns that do not.
- After learning is applied, MaxEnt does not severely overgenerate.

Patterns Under Investigation

There are two types of variation predicted by MaxEnt:

- **Normal Variation**- Most of the probability is split between candidates that could surface categorically.

Variable Onset Epenthesis

$$/\text{pa}/ \quad \rightarrow \quad [\text{pa}] \quad 100\% \qquad \boxed{/\text{u}/ \quad \rightarrow \quad \begin{array}{ll} [\text{u}] & 50\% \\ [\text{ʔu}] & 50\% \end{array}}$$

- **Harmonically-Bounded Variation**- Most of the probability is split between candidates some of which are harmonically bounded.

Variable Onset Deletion

$$\boxed{/\text{pa}/ \quad \rightarrow \quad \begin{array}{ll} [\text{pa}] & 50\% \\ [\text{a}] & 50\% \end{array}} \qquad /\text{u}/ \quad \rightarrow \quad [\text{u}] \quad 100\%$$

Solution

I present learning simulations that show:

- Harmonically-bounded variation takes more data to learn than normal variation.
- Harmonically-bounded variation is less stably transmitted across generations.
- If both normal and harmonically bounded variation are found in a pattern, the harmonically-bounded variation will be lost first.

When filtered by learnability, MaxEnt is unlikely to give (much) probability to harmonically bounded candidates.

Relevant weighting condition

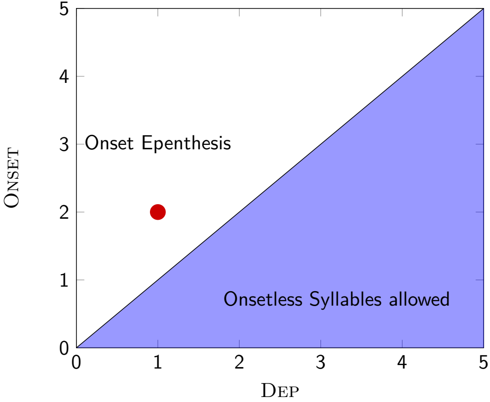
In Categorical HG:

- Onsets epenthesize when ONSET outweighs DEP.
- Onsetless syllables remain faithful when DEP outweighs ONSET.

In MaxEnt:

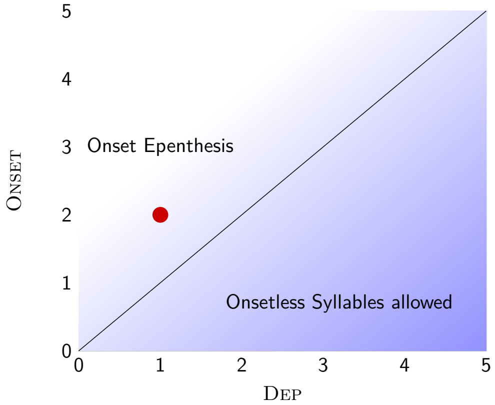
- Onsets epenthesize *more* when ONSET outweighs DEP *more*.
- Onsetless syllables remain faithful *more* when DEP outweighs ONSET *more*.

Categorical Harmonic Grammar



V	ONSET	DEP	
Weights	$w = 2$	$w = 1$	HARMONY
a. V	-1		-2
☞ b. CV		-1	-1

MaxEnt



/V/	ONSET	DEP		
Weights	$w = 2$	$w = 1$	HARM	PROB
a. V	-1		-2	.27
b. CV		-1	-1	.73

Relavant Constraint Weighting

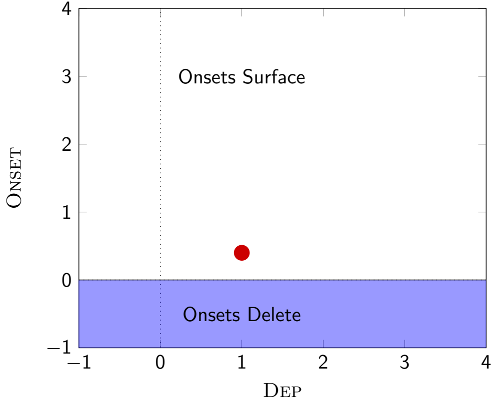
In Categorical HG:

- Onsets delete when MAX + ONSET is lower than zero.
- Onsets are preserved when the sum of MAX + ONSET is above zero.

In MaxEnt:

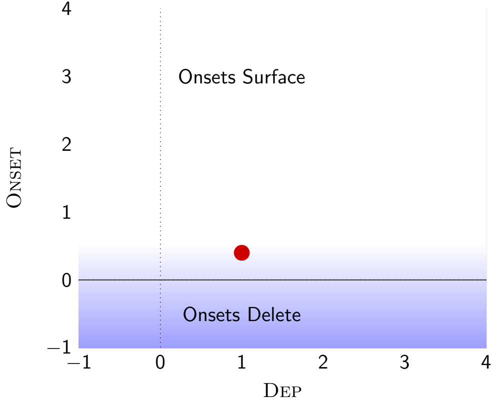
- Onsets delete *more* when MAX + ONSET is lower.
- Onsets are preserved *more* when the sum of MAX + ONSET is higher.

Harmonic Bounding - Categorical HG



/CV/	ONSET	DEP	
Weights	$w = .4$	$w = 1$	HARMONY
a. CV			0
☞ b. V	-1		-.4

Harmonic Bounding - MaxEnt



/CV/	ONSET	DEP		
Weights	$w = .4$	$w = 1$	HARM	PROB
a. CV			0	.60
b. V	-1		-.4	.40

Modeling Learning X Grammar

- Learning bias is a diachronic pressure
- A child may not end up learning the same grammar as their parent
  - If the parent's target grammar is harder to learn, the learner has higher probability of mislearning
- Even a small probability of mislearning can have a large effect on typology over many generations.
- By modeling generational transmission (Kirby & Huford, 2002; Staubs, 2014; Kirby, 2017; Hughto, 2018; O'Hara, in prep), we can observe effect of learning bias.

Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected (here, a syllable structure).
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher's form more likely in the future.

Teacher		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	50%
	[V]	50%

Learner		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	75%
	[V]	25%

Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected (here, a syllable structure).
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher's form more likely in the future.

Teacher		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	50%
	[V]	50%

/V/

Learner		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	75%
	[V]	25%

Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected.
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher's form more likely in the future.

Teacher		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	50%
	[V]	50%

/V/

[V] [CV]

Learner		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	75%
	[V]	25%

Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected.
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher's form more likely in the future.

Teacher		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	50%
	[V]	50%

/V/

[V] ✗ [CV]

Learner		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	75%
	[V]	25%

Simulation Methodology: Within-Generation

I use the truncated perceptron algorithm Magri (2015); Rosenblatt (1958); Boersma & Pater (2016).

- An input is randomly selected.
- The teacher and learner select outputs for that input based on their current grammar.
- If they differ, the learner updates their grammar to make the teacher's form more likely in the future.

Teacher		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	50%
	[V]	50%

/V/

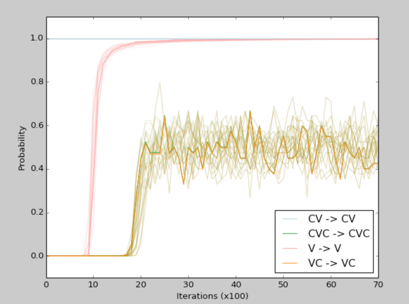
[V] ✗ [CV]

Learner		
/CV/	[CV]	100%
	[V]	0%
/V/	[CV]	70%
	[V]	30%

Two Phases of Error Driven Learning

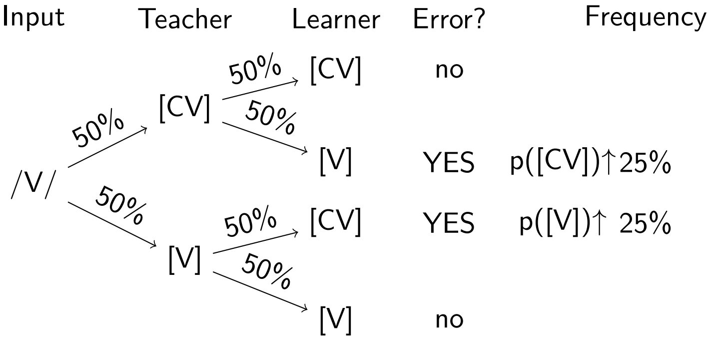
There are two major phases of error-driven learning of stochastic grammars.

- **Learning Phase**: Most updates move the learner towards the target grammar and away from the starting grammar.
- **Oscillation Phase**: Updates cause the learner to oscillate around the target pattern.



Oscillatory Phase

When the teacher's grammar is variable, errors continue even when the learner has the same grammar.



- Errors occur 50% of the time, but they are balanced in both directions, so the average across many runs will remain at the target pattern.

Simulation Methodology: Across-Generation

Generational Learning Model[4]

- Simulated learners using MaxEnt grammars
- Learners are initialized with Markedness constraints high, faith low[5]
- Train a learning agent off of some limited number of forms[6] from a teacher.
- Then train a new learner on that agent's final grammar.
- Patterns that remain stable across generations are likely to be better attested.

[4] Following Staubs (2014); Hughto (2018)

[5] Gnanadesikan (2004); Tesar & Smolensky (2000); Jesney & Tessier (2011)

[6] Kirby & Huford (2002)

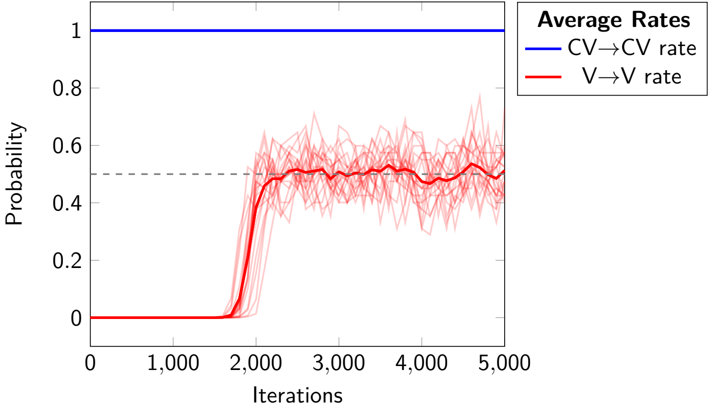
NORMAL VARIATION SIMULATIONS

Variable Onset Epenthesis

/pa/	→	[pa]	100%	/u/	→	[u]	50%
						[ʔu]	50%

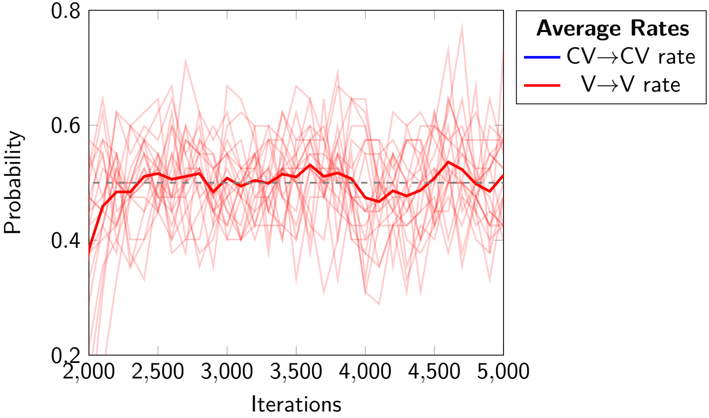
Learning of Normal Variation

Normal variation simulations clearly show the oscillation phase, but the average run converges towards the target grammar.



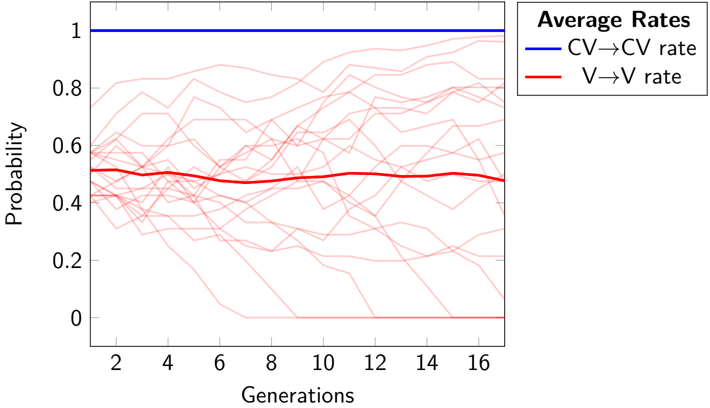
Learning of Normal Variation

Normal variation is learned here in around 2100 iterations.



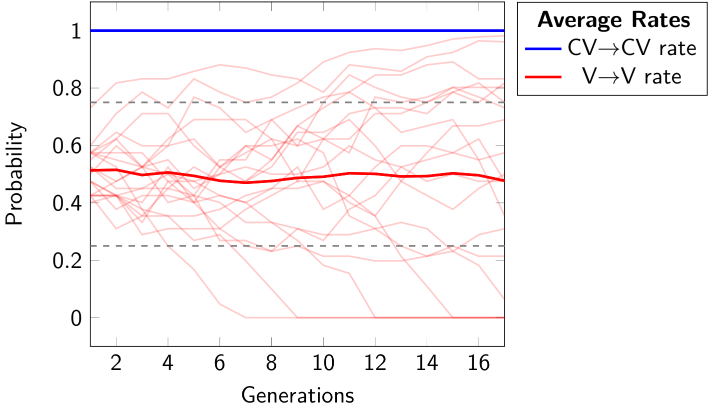
Generational Change of Normal Variation

There is variation across runs in terms of generational change.
Typological implications are respected in all runs throughout.



Generational Change of Normal Variation

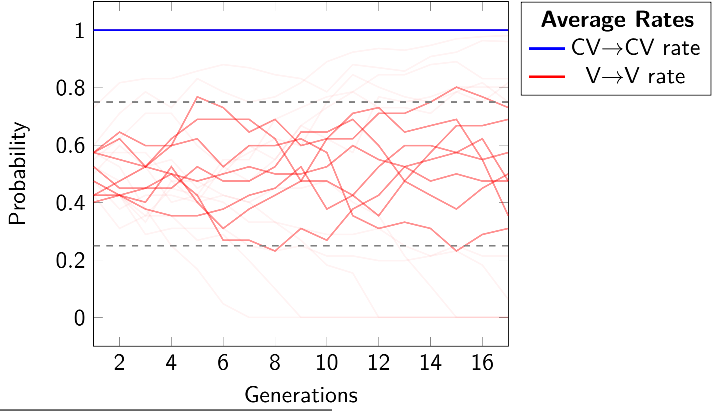
40% remained within a .25 probability window, but 12 runs lost variation: six categorically epenthesize onsets, and six never epenthesize onsets. [7]



[7] bias for categorical patterns replicating Hughto (2018)

Generational Change of Normal Variation

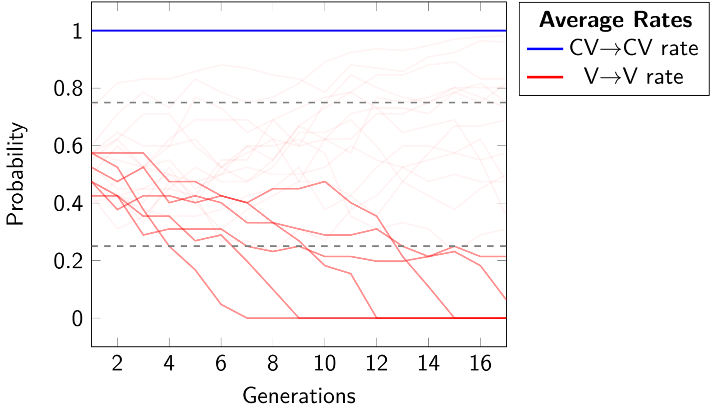
40% remained within a .25 probability window, but 12 runs lost variation: six categorically epenthesize onsets, and six never epenthesize onsets. [8]



[8] bias for categorical patterns replicating Hughto (2018)

Generational Change of Normal Variation

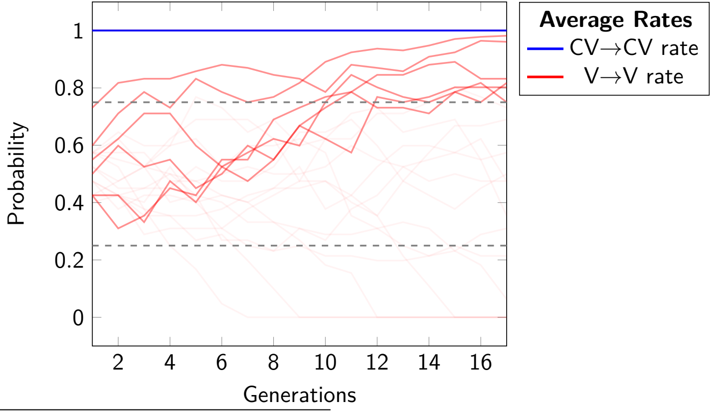
40% remained within a .25 probability window, but 12 runs lost variation: **six categorically epenthesize onsets**, and six never epenthesize onsets. [9]



[9] bias for categorical patterns replicating Hughto (2018)

Generational Change of Normal Variation

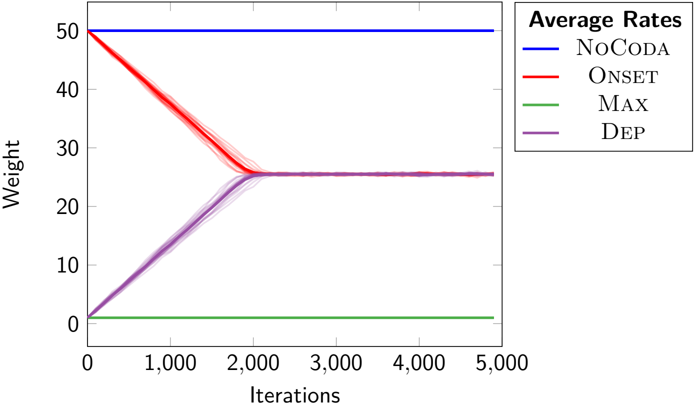
40% remained within a .25 probability window, but 12 runs lost variation: six categorically epenthesize onsets, and **six never epenthesize onsets.** [10]



[10] bias for categorical patterns replicating Hughto (2018)

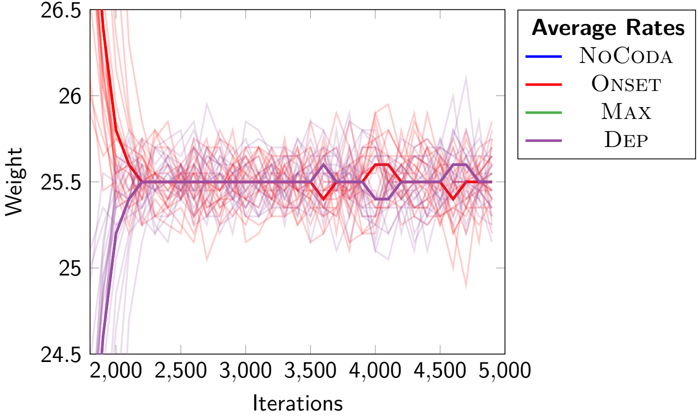
Weights for Normal Variation

First generation weighting dynamics are consistent, ONSET and DEP meet



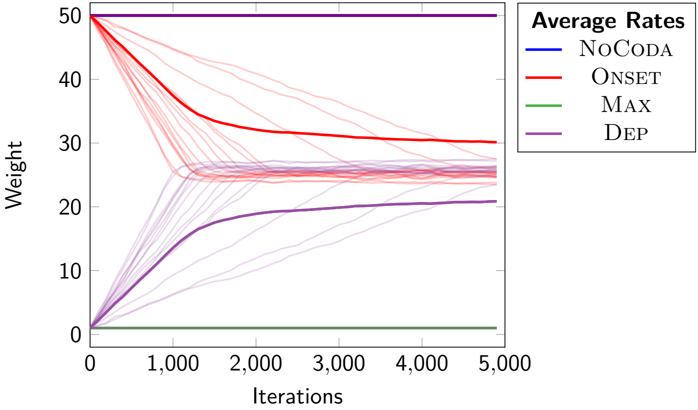
Weights for Normal Variation

And then they oscillate around eachother.



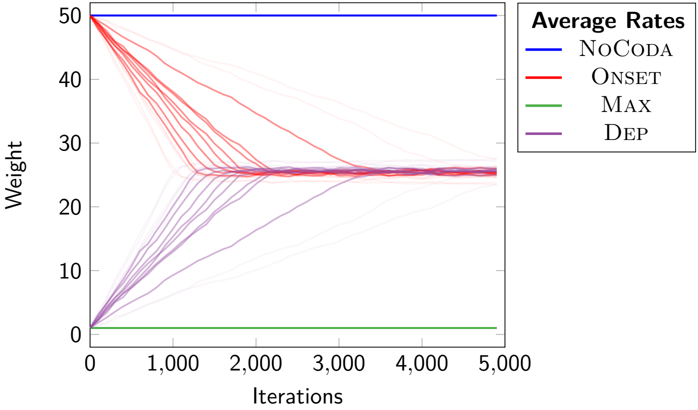
Weights at later generations

At the seventeenth (last) generation, there are 3 types of weighting dynamics observed.



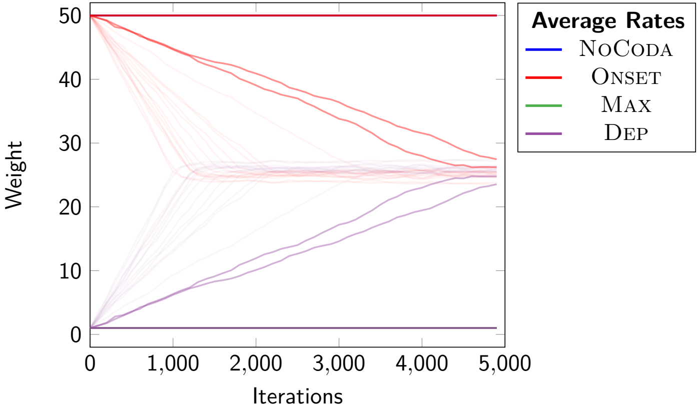
Weights at later generations

At the seventeenth (last) generation, there are 3 types of weighting dynamics observed. **Variation**



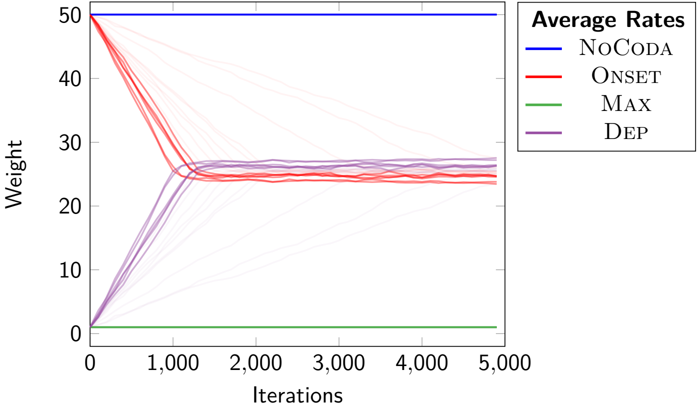
Weights at later generations

At the seventeenth (last) generation, there are 3 types of weighting dynamics observed. Variation, **Categorical Faithfulness**



Weights at later generations

At the seventeenth (last) generation, there are 3 types of weighting dynamics observed. Variation, Categorical Faithfulness, and **Categorical Epenthesis**



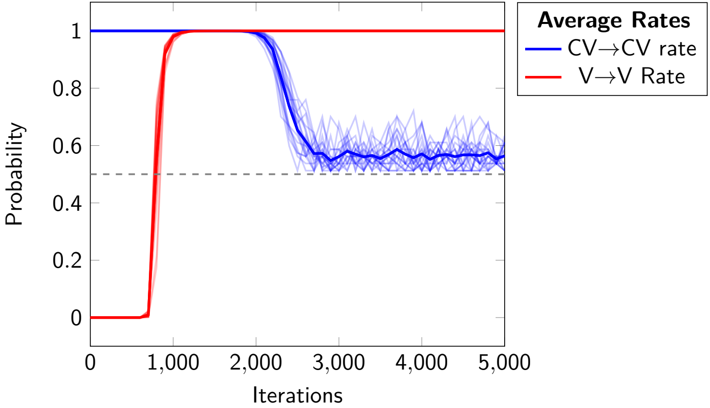
HARMONICALLY-BOUNDED VARIATION SIMULATIONS

Variable Onset Deletion

$\boxed{/\text{pa}/ \rightarrow \begin{matrix}[\text{pa}] & 50\% \\ [\text{a}] & 50\%\end{matrix}}$	/u/	$\rightarrow$	[u]	100%

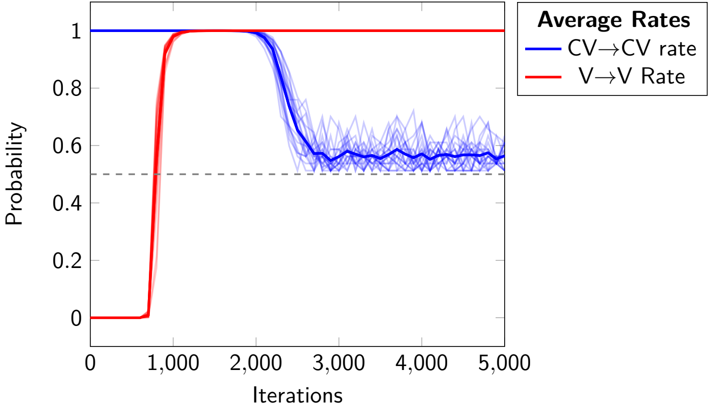
Learning of Harmonically-Bounded Variation

Given enough data, learners can learn to delete onsets.



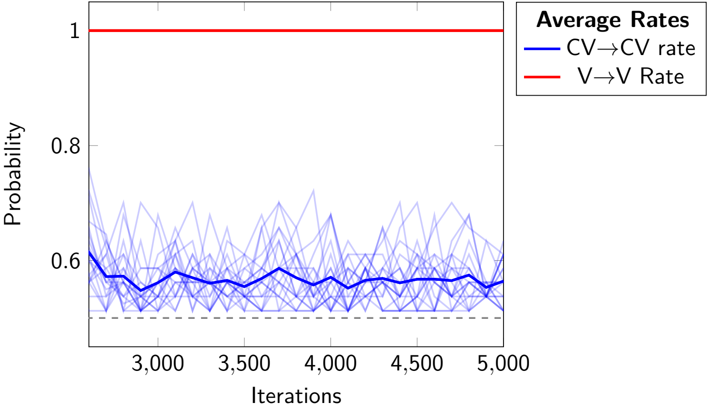
Learning of Harmonically-Bounded Variation

Given enough data, learners can learn to delete onsets. Converges at around 2700 iterations.



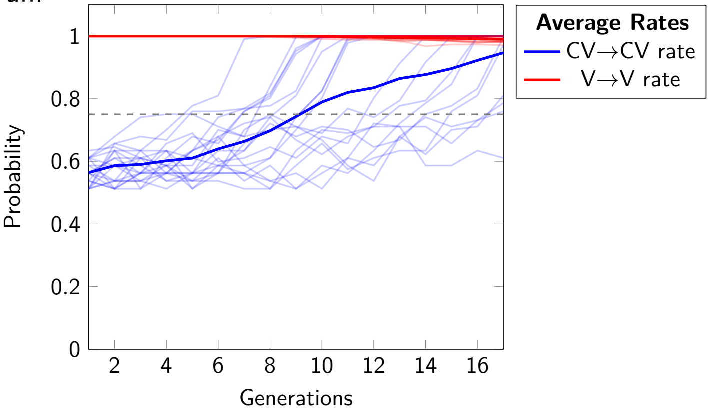
Learning of Harmonically-Bounded Variation

But notably, it doesn't converge quite to the target pattern (gray dashed line).



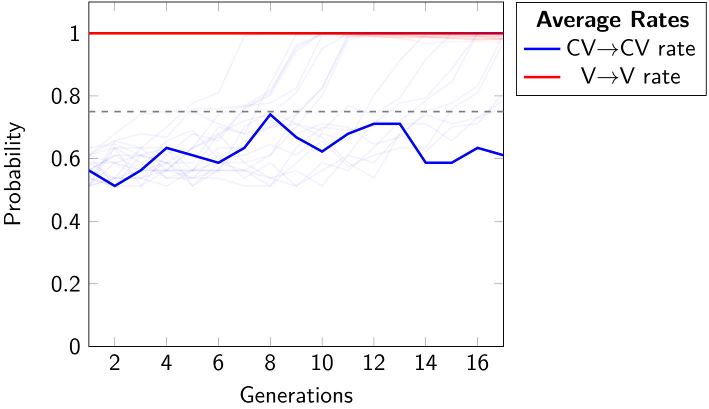
Generational Change of Harmonically-Bounded Variation

Harmonically-Bounded Variation is far less stable—lost in all but one run.



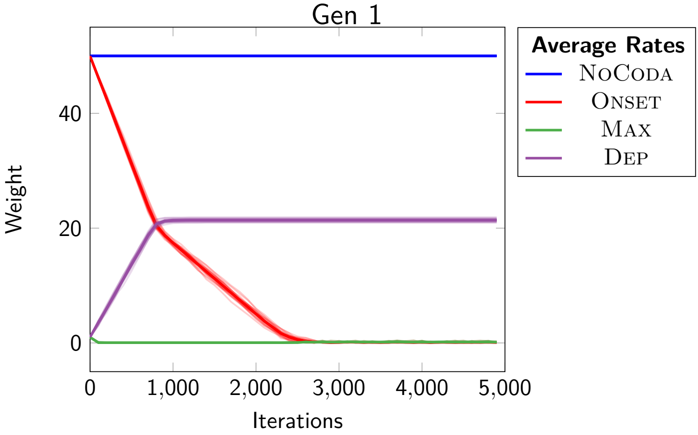
Generational Change of Harmonically-Bounded Variation

Harmonically-Bounded Variation is far less stable—lost in all but **one run.**



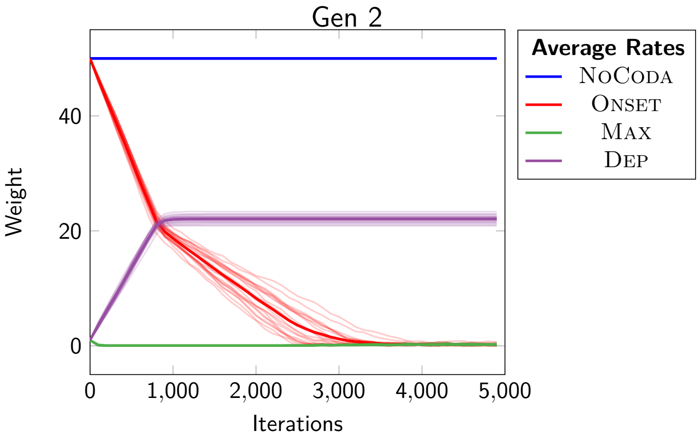
Weights for Harmonically-Bounded Variation

First generation weighting dynamics are consistent, constraints need to go to zero so harmonic bounded candidates get weight



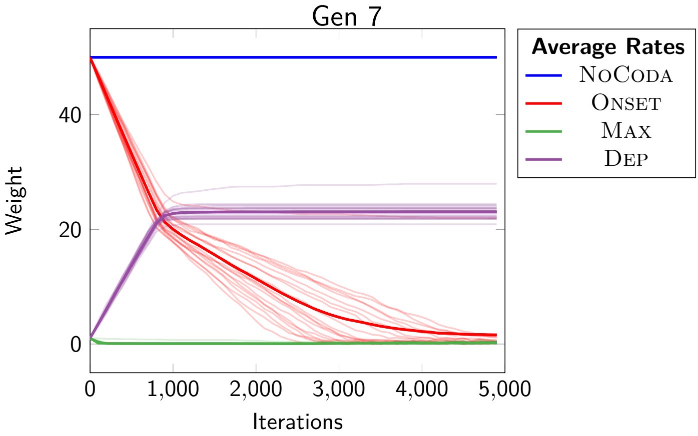
Weights for Harmonically-Bounded Variation

In later generations, it takes increasingly long for ONSET to reach zero.



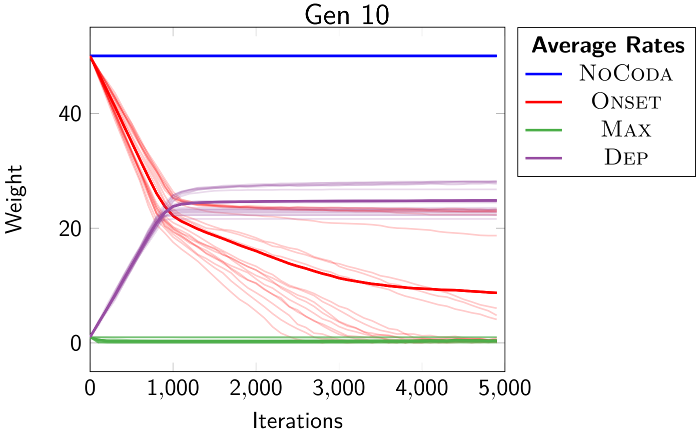
Weights for Harmonically-Bounded Variation

In later generations, it takes increasingly long for ONSET to reach zero.



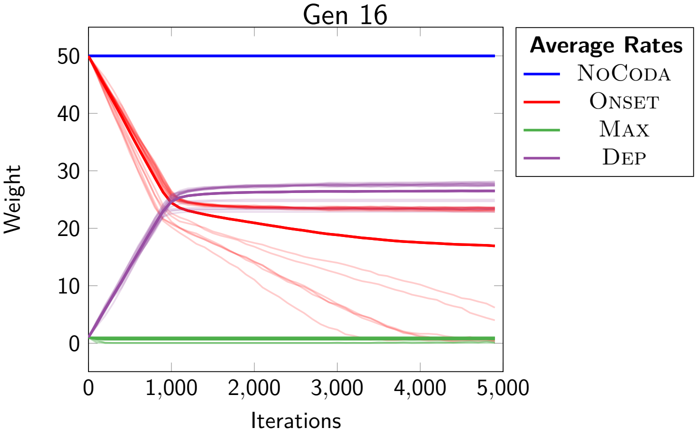
Weights for Harmonically-Bounded Variation

Once a learner doesn't reach near zero for ONSET, they quickly stop lowering it far below DEP



Weights for Harmonically-Bounded Variation

Once a learner doesn't reach near zero for ONSET, they quickly stop lowering it far below DEP



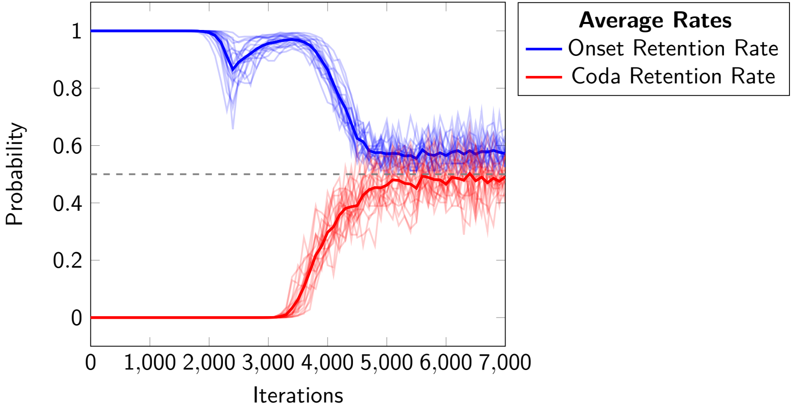
BOTH TYPES OF VARIATION SIMULATIONS

Variable Onset and Coda Deletion

/pa/	$\rightarrow$	[pa] [a]	50% 50%	/u/	$\rightarrow$	[u]	100%
/a/	$\rightarrow$	[a]	100%	/uk/	$\rightarrow$	[uk] [u]	50% 50%

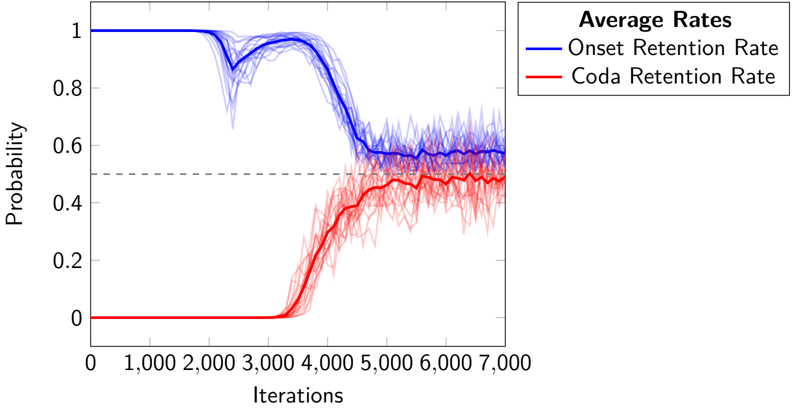
Learning of Harmonically-Bounded Variation

What if we look at both types of variation in one grammar?



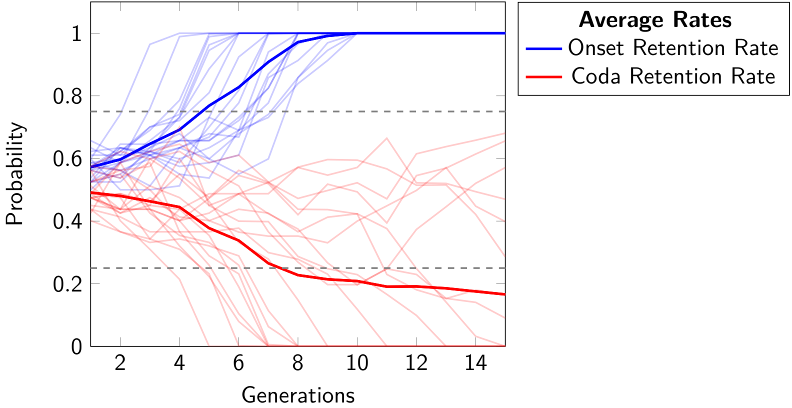
Learning of Harmonically-Bounded Variation

What if we look at both types of variation in one grammar?
Converges at around 5000 iterations.



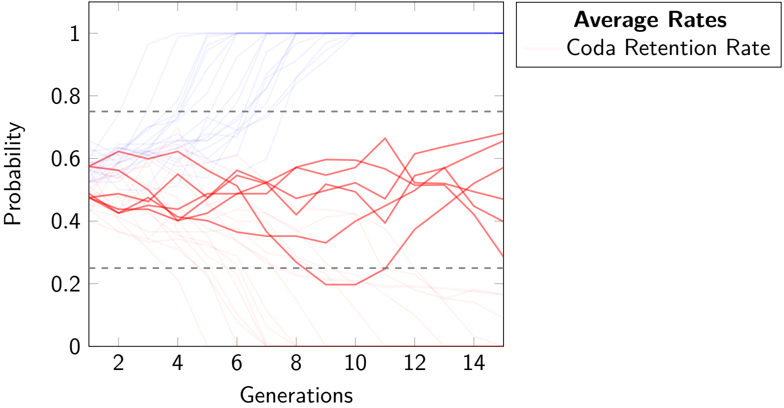
Generational Change of Combined Variation

Harmonically-Bounded Variation is far less stable—all runs lose harmonically bounded variation.



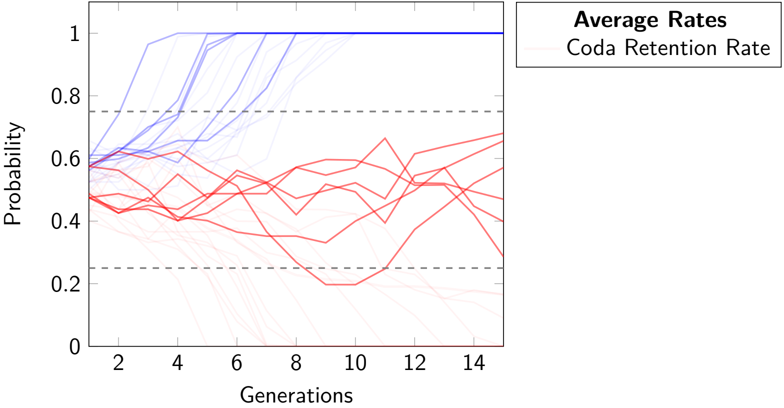
Generational Change of Combined Variation

Harmonically-Bounded Variation is far less stable—all runs lose harmonically bounded variation but 30% maintain normal variation.



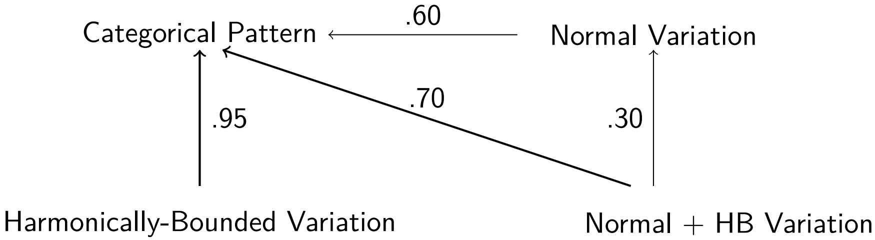
Generational Change of Combined Variation

Harmonically-Bounded Variation is far less stable—all runs lose harmonically bounded variation but 30% maintain normal variation.



Summary

- Harmonically Bounded Variation is harder to learn than Normal Variation.
- If Harmonically Bounded Variation and Normal Variation are in a pattern, loss of Normal Variation implies loss of Harmonically Bounded Variation.



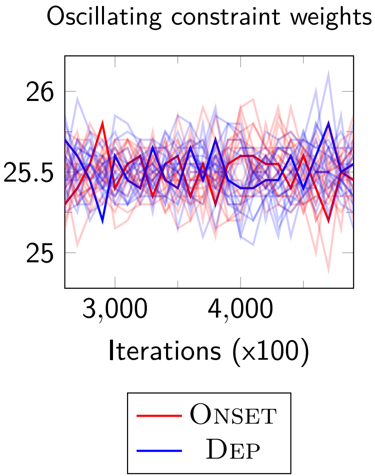
Why is giving probability to harmonically bounded candidates hard?

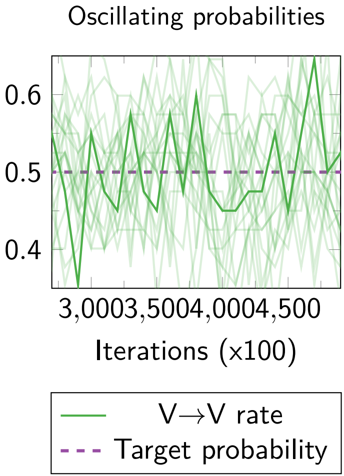
- Different types of weighting condition needed to give harmonically bounded candidate probability.

Normal Variation	**HB Variation**
DEP ∼ ONSET	MAX ∼ ONSET∼0
2100 Iterations	2700 Iterations

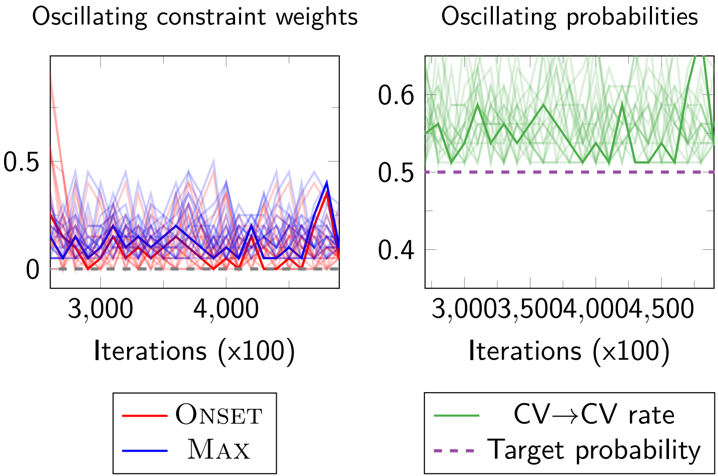
- Oscillation phase works different for harmonically-bounded variation.
  - In a normal variation pattern, the learner is equally likely to oscillate towards either candidate.
  - In a harmonically-bounded variation pattern, the truncated aspect of the algorithm bounds how much probability the learner can give the harmonically bounded candidate.

Generational Differences





Generational Differences



Harmonically Bounded Candidates

With learning, MaxEnt rarely gives harmonically bounded candidates much probability, but it always will give them SOME probability.
Giving some probability to harmonically bounded candidates may not be the worst thing in the world.

- A harmonically bounded candidate never receives the most probability of the candidates.
- Harmonically bounded candidates can be observed in speech errors[11]
- Gradient well-formedness of harmonically bounded candidates can be greater than non-bounded candidates [12]

[11] Goldrick & Daland (2009)

[12] Hayes & Wilson (2008); Hayes & Moore-Cantwell (2011); Hayes (2017); Hayes & Schuh (to appear)

Takeaway

- Is grammatical overgeneration a problem?
  - Not necessarily, if the unattested languages can be ruled out independently by learning (or other factors)
- Does grammatical structure still matter?
  - Yes! Properties of the grammar (like harmonic bounding) still have some effect.

Takeaway

- Is grammatical overgeneration a problem?
  - Not necessarily, if the unattested languages can be ruled out independently by learning (or other factors)
- Does grammatical structure still matter?
  - Yes! Properties of the grammar (like harmonic bounding) still have some effect.

Takeaway

- Is grammatical overgeneration a problem?
  - Not necessarily, if the unattested languages can be ruled out independently by learning (or other factors)
- Does grammatical structure still matter?
  - Yes! Properties of the grammar (like harmonic bounding) still have some effect.

Takeaway

- Is grammatical overgeneration a problem?
  - Not necessarily, if the unattested languages can be ruled out independently by learning (or other factors)
- Does grammatical structure still matter?
  - Yes! Properties of the grammar (like harmonic bounding) still have some effect.

Works Cited I

ANTTILA, ARTO, & MAGRI, GIORGIO. 2018. Does MaxEnt overgenerate? Implicational universals in Maximum Entropy grammar. *In:* GALLAGHER, GILLIAN, GOUSKOVA, MARIA, & SORA, YIN (eds), *Proceedings of the 2017 Annual Meeting on Phonology*. Linguistics Society of America.

BOERSMA, PAUL. 2003. Review of Bruce Tesar and Paul Smolensky 2000, Learnability in Optimality Theory. *Phonology*, **20**, 436–446.

BOERSMA, PAUL, & PATER, JOE. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. *In:* MCCARTHY, JOHN J., & PATER, JOE (eds), *Harmonic Grammar and Harmonic Serialism*. Equinox.

GNANADESIKAN, AMALIA. 2004. Markedness and Faithfulness in child phonology [ROA-67]. *Pages 73–108 of:* KAGER, RENÉ, PATER, JOE, & ZONNEVELD, WIM (eds), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press.

GOLDRICK, MATT, & DALAND, ROBERT. 2009. Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology*, **26**, 147–185.

HAYES, BRUCE. 2017. Varieties of Noisy Harmonic Grammar. *In:* JESNEY, KAREN, O'HARA, CHARLIE, SMITH, CAITLIN, & WALKER, RACHEL (eds), *Proceedings of the 2016 Annual Meeting on Phonology*. Washington, DC: Linguistics Society of America.

HAYES, BRUCE, & MOORE-CANTWELL, CLAIRE. 2011. Gerard Manley Hopkin's sprung rhythm: corpus study and stochastic grammar. *Phonology*, **28**, 235–282.

HAYES, BRUCE, & SCHUH, RUSSELL G. to appear. Metrical structure and sung rhythm of the Hausa rajaz. *Phonological Analysis*.

HAYES, BRUCE, & WILSON, COLIN. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, **39**, 379–440.

HUGHTO, CORAL. 2018. Investigating the Consequences of Iterated Learning in Phonological Typology. *In:* *Proceedings of the Society for Computation in Linguistics*, vol. 1.

Works Cited II

JESNEY, KAREN. 2007. *The locus of variation in weighted constraint grammars*. Handout for poster presented at the Workshop on Variation, Gradience and Frequency in Phonology, Stanford University.

JESNEY, KAREN, & TESSIER, ANNE-MICHELLE. 2011. Biases in Harmonic Grammar: The road to restrictive learning. *Natural Language & Linguistic Theory*, **29**.

KIRBY, SIMON. 2017. Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin and Review*, **24**, 118–137.

KIRBY, SIMON, & HUFORD, JAMES. 2002. The emergence of linguistic structure: An overview of the iterated learning model. *Chap. 6, pages 121–148 of:* CANGELOSI, A, & PARISI, D. (eds), *Simulating the Evolution of Language*. London: Springer Verlag.

LEGENDRE, GÉRALDINE, MIYATA, YOSHIRO, & SMOLENSKY, PAUL. 1990. Harmonic Grammar - a formal multi-level conectionist theory of linguistic wellformedness: an application. *Pages 884–891 of:* ERLBAUM, LAWRENCE (ed), *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*.

MAGRI, GIORGIO. 2015. How to keep the HG weights non-negative: the truncated Perceptron reweighting rule. *Journal of Language Modeling*, **3**(2), 345–375.

MCCARTHY, JOHN J., & PRINCE, ALAN. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers*, **18**, 249–384.

O'HARA, CHARLIE. 2017. How abstract is too abstract: Learning abstract underlying representations. *Phonology*, **34**(2), 325–345.

O'HARA, CHARLIE. 2018. *Learnability Captures Soft Typology of Coda Stop Inventories,*. Presented at LSA 2018.

O'HARA, CHARLIE. 2019. *Emergent Learning Bias and the Underattestation of Simple Patterns*. Ms. University of Southern California.

O'HARA, CHARLIE. in prep. *Soft Typology in Phonology: Learnability meets grammar*. Ph.D. thesis, University of Southern California.

Works Cited III

PATER, JOE. 2016. Universal Grammar with Weighted Constraints. *Pages 1–46 of:* MCCARTHY, JOHN J., & PATER, JOE (eds), *Harmonic Grammar and Harmonic Serialism*. London: Equinox.

PATER, JOE, & MORETON, ELLIOTT. 2012. Structurally biased phonology: complexity in language learning and typology. *The EFL Journal*, **3**(2), 1–44.

PRINCE, ALAN, & SMOLENSKY, PAUL. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

ROSENBLATT, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**, 386–408.

STAUBS, ROBERT. 2014. *Computational modeling of learning biases in stress typology*. Ph.D. thesis, University of Massachusetts Amherst, Amherst.

TESAR, BRUCE, & SMOLENSKY, PAUL. 2000. *Learnability in Optimality Theory*. MIT Press.

Questions

I focused here on simply harmonically bounded forms. Collectively bounded forms may act different.

/bababa/	*B	FAITH		
weights	$w = 5$	$w = 5$	HARM	PROB
a. bababa	-3		-15	.33
b. bapaba	-2	-1	-15	.33
c. papapa		-3	-15	.33

- Noisy HG performs differently than MaxEnt here (Hayes, 2017).
  - The version discussed in most of this paper gives no probability to the bounded candidate.
  - Other versions can create a u-shaped distribution across these forms.
- Can these types of patterns cause subversion of t-orders?
- Can the distribution of probability across collectively bounded forms (local optionality) differentiate between theories? Maybe (Hayes, 2017)

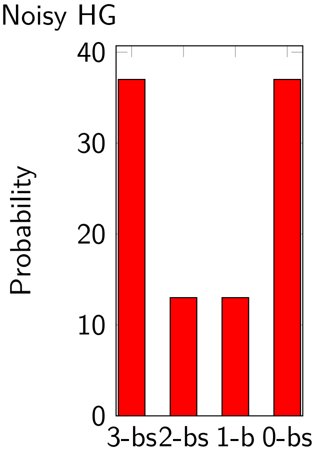
Questions

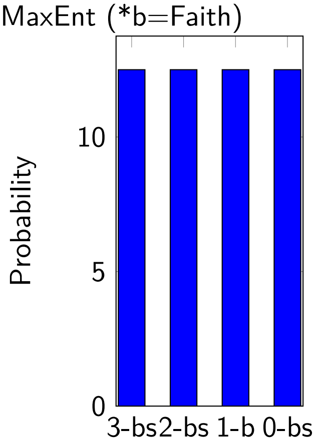
I focused here on simply harmonically bounded forms. Collectively bounded forms may act different.

/bababa/	*B	FAITH		
weights	$w = 5.1$	$w = 5$	HARM	PROB
a. bababa	-3		-15.3	.33
b. bapaba	-2	-1	-15.2	.33
c. papapa		-3	-15	.33

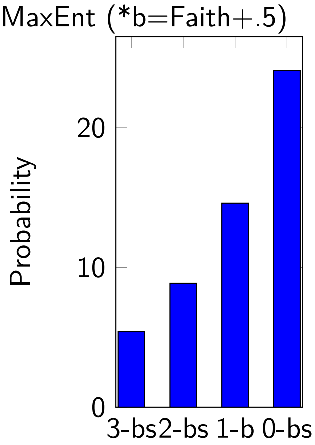
- Noisy HG performs differently than MaxEnt here (Hayes, 2017).
  - The version discussed in most of this paper gives no probability to the bounded candidate.
  - Other versions can create a u-shaped distribution across these forms.
- Can these types of patterns cause subversion of t-orders?
- Can the distribution of probability across collectively bounded forms (local optionality) differentiate between theories? Maybe (Hayes, 2017)

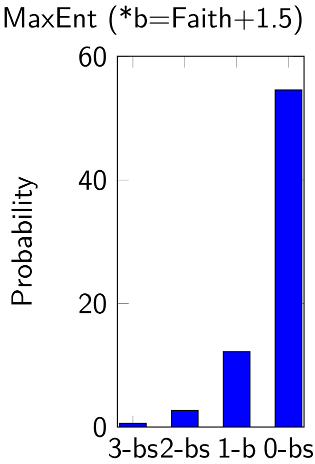
Collectively Bounded Forms





Collectively Bounded Forms





Do learners see harmonically bounded forms?

In these simulations, learners had full access to the teacher's underlying form.

- This is unlikely from a learning standpoint.
- If a teacher produces /CVC/-[VC], a learner only hears [VC].
- By Lexicon Optimization, the learner will usually choose /VC/ as the underlying representation, rather than the harmonically bounded /CVC/.
- Harmonically bounded mappings are all either unfaithful, or involve hidden structure.
- Thus, learners would perceive even fewer harmonically bounded mappings than in these simulations.