Maximal feature specification is feasible; minimal feature specification is not

Kyle Gorman
CUNY Graduate Center

Charles Reiss
Concordia University

GLOW 46, Vienna

Three heresies

- Phonological rules are not minimally specified. Rather they are specified using the **maximal set of features** consistent with the data.
- **Children are not little scientists**; acquiring a grammar is qualitatively different than developing a scientific theory.
- **There's no such thing** as assimilation.

1 Our assumptions

- There is an innate, finite, and universal set of bivalent[1] phonological features (e.g., Chomsky and Halle 1965, Reiss and Volenec 2022).
- Segments are sets of specified features.
- Natural classes are sets of sets of specified features defined by generalized intersection.
- Rules are defined with natural classes ("no disjunction", "no output constraints", apparent exceptions necessarily reflect multiple rules and possibly, rule ordering).
- The acquirer makes generalizations licensed by the input ("no negative evidence").
- There is no language-specific phonetic implementation (e.g., Chomsky and Halle 1965, Reiss and Volenec 2022).

2 Where am I?

- Do kids assume they are in Reykjavik or Rochester?

(1)
a. Sigga$_i$ segir að Maria$_j$ elski sig$_{i/j}$ (Icelandic)
b. Sigga$_i$ says that Maria$_j$ loves herself$_{*i/j}$ (English)

[1]This assumption does not lead to any loss of generality, but we retain it for ease of exposition.

(2) English ⊆ Icelandic

- A straightforward application of the Subset Principle requires that kids initially assume that anaphors are locally bound, as in English.
- Positive evidence allows them to relax that restriction and acquire Icelandic.

- Do kids assume they are in Riyadh or Reynosa?

(3)	a.	/i, a, u/	(Arabic)
	b.	/i, e, a, o, u/	(Spanish)

(4) Spanish ⊆ Arabic

- The Subset Principle requires kids to assume a maximally specified vowel inventory, assuming inventories are defined intensionally (e.g., Hale and Reiss 2003).
- Each vowel initially corresponds to a highly restricted (highly specified) region of vowel space; e.g., Spanish /i, e/ are subsets (i.e., subspaces) of Arabic /i/.
- A subspace is **more** specified: [−Back, −High] is a subspace of [−Back].
- Positive evidence (e.g., that ±High is non-contrastive) allows them to "prune" that distinction and acquire Arabic.

3 Georgian

- Georgian has two laterals in complementary distribution: plain/light [l] occurs before the front vowels [i, e], and the velarized/dark [ɫ] occurs elsewhere.

(5) Georgian laterals:

a.	**l**e**ɫ**o	'goal'	(Robins and Waterson 1952)
	k'b̥i**ɫ**s	'tooth (dat.)'	
	tsʰo**l**i	'wife'	
	ɫamazad̥	'prettily'	
	ɑ**ɫ**q'ɑ	'siege'	
b.	go**l**i	'goal'	("Natia", Manhattan-bound Q train, 3/21/23)
	k'b̥i**l**i	'tooth (nom.)'	
	ɫamazi	'pretty'	
	ɫudi	'beer'	

- One might write this rule informally as follows:

(6) ɫ → l / __ /i, e/

- If we wish to specify (6) formally, how general or specific should the environment be?

(7)
a. __ [−Back]
b. __ [−Back, −Low]
c. __ [−Back, −Low, −Round, −Nasal, −Creaky, +Voice, . . .]

- **There is no empirical way to decide** on the basis of Georgian alone, because Georgian −Back vowels are −Low, −Round, and −Nasal,[2]. so the specifications in (7) are all **extensionally equivalent**.
- A similar point can be made for *target* specificity:

(8)
a. [+Lateral] → {−Back}
b. [+Lateral, +Sonorant, −Continuant, +Voice, . . .] → {−Back}

- Feature specificity question isn't a purely philosophical matter: **both linguists and infants acquiring Georgian require a satisfactory answer**.

4 The received wisdom

- Most phonologists assume rules are specified "minimally", with few features as possible.
- *The Sound Pattern of English* (Chomsky and Halle 1968:§8.1) proposes a feature-counting evaluation metric which favors the most concise empirically adequate grammar.

> Conciseness Condition: If there is more than one possible grammar that can be constructed for a given body of data, choose the grammar that is most concise in terms of the number of feature specifications. (Kenstowicz and Kisseberth 1979:336)

- This is still received wisdom in modern phonology textbooks:

> There are good reasons to include only just as many features in a rule as are needed. (Hayes 2009:92)

> ...one should use the minimum number of features required to specify all and only the sounds in the class. (Zsiga 2012:282)

> ...rules are stated in terms of the simplest, most general classes of phonetically defined segments... (Odden 2013:66–67)

- Odden justifies minimization with reference to Occam's Razor. He writes:

[2]We put aside the question of whether features which are non-constrastive in a language—like Creaky in Georgian—are visible to the phonology. While this is an interesting problem, the specificity issue remains if we focus on comparing (7ab) and put aside (7c), since Back and Low are clearly contrastive in this language. See Hale and Reiss (2008:§2.6) for some discussion.

> In this example we only have direct evidence for the change after *m*, so it would be possible to restrict our rule to the more specific context "after *m*." But this would run counter to basic assumptions of science, that we seek the most general explanations possible, not the most restricted. (*ibid.*:89)

- Similar points are made by Dell and by Spencer:

> Between two grammars that generate the same set of descriptions of sound-meaning pairs, the one containing the smallest number of rules, and the rules that have the most general scope, is chosen; in other words, the simplest grammar. This is an absolutely fundamental point. All linguists proceed in this manner, whatever school they belong to, and if they are often not aware of it, it is generally because they are not aware that for any set of data, their theoretical presuppositions (whether they are explicit or not) allow for a great number of competing descriptions. (Dell 1980:138)

> Linguists, like other scientists, like to provide the most general statement of a rule or a principle. (Spencer 1996:136)

- These statements conflate *scientific epistemology*—the techniques by which scientists painstakingly discover truths about the world—with child language acquisition.
  - Scientists don't know much: they need Occam's razor.
  - Infants have Universal Grammar: they don't need heuristics.
- Recent work (e.g., Rasin et al. 2021:17) revives the minimal specification approach using the information-theoretic notion of *minimum description length*.

5 Our proposal

- Appeals to minimal specification are largely informal: they propose an objective, not an algorithm. When we attempt to develop this algorithm, we run into problems.
- We first present arguments against minimal specification:
  - Minimal specification is infeasible (i.e., computationally intractable).
  - Minimal specification is not unique (i.e., it need not have a single solution).
- Then, we propose **maximal specification** as a simple, feasible, and unique alternative.

6 Feasibility

- How does one pick the minimal, empirically adequate natural class from among the 3^n intensionally distinct natural classes (where n is the number of features)?
  - For $n = 5$, $3^n = 243$.
  - For $n = 24$, $3^n \approx$ 282 billion (2.8243×10^{11}).

- More generally, **how hard is this problem?**
- An algorithm is said to run in polynomial time if it is guaranteed to halt in time proportional to some polynomial of the size of the input.
- According to the *P-Cognition thesis* (e.g., Frixione 2001), computational models of cognition are feasible only if they can be solved in polynomial time.

> There is wide agreement that a problem has not been "well-solved" until a polynomial time algorithm is known for it. Hence, we shall refer to a problem as intractable, if it is so hard that no polynomial time algorithm can possibly solve it. (Garey and Johnson 1979:8; quoted in van Rooij 2008)
>
> ...cognitive capacities are limited to those functions that can be computed in polynomial time (van Rooij 2008:948).

- The P-Cognition thesis has been—implicitly—adopted by computational phonologists:
  - Eisner (1997), Idsardi (2006), and Heinz et al. (2009) debate whether it is possible to find the optimal candidate in Optimality Theory in polynomial time.
  - Heinz (2010) proves his algorithm for acquiring long-distance phonotactic generalizations is polynomial time.
  - Chandlee et al. (2014) emphasize that their algorithm for acquiring phonological mappings is polynomial time.
- One can prove that there is no polynomial time algorithm by reducing it from (sic) a previously studied problem which itself cannot be solved in polynomial time.
- Chen and Hulden (2018, henceforth C&H) prove that feature minimization is as difficult as the problem known as *set cover*, which is known to be *NP-complete* (Karp 1972).[3]
- Under the standard conjecture (that $P \neq NP$), **NP-complete problems—including feature minimization**—cannot be solved by any algorithm in polynomial time, so they **are not feasible cognitive models** according to the P-Cognition thesis.
- C&H also experiment—unsuccessfully—with heuristics for feature minimization.
  - *Greedy search* fails to find minimal specifications, even when the set in question is a single phoneme (e.g., /ə/).
  - *Branch-and-bound* algorithms are usually able to find a minimal specification, but may still need to search hundreds of thousands of possible specifications.

[3] We assume C&H's results, but omit their reduction for reasons of space and time; see their study for details.

7 Uniqueness

- Georgian has a five-vowel inventory:

> Phonologically, *i* and *e* may be designated as front vowels, *a*, *o*, and *u* as back vowels. (Robins and Waterson 1952:59)

(9) Georgian vowel features:

	High	Low	Back
i	+	−	−
e	−	−	−
a	−	+	+
o	−	−	+
u	+	−	+

- Using the contrastive features in (9), one can construct:
  - 21 *intensionally distinct* non-empty natural classes, corresponding to
  - 13 *extensionally distinct* non-empty natural classes.
- One might suppose minimization provides a sensible way to decide between intensionally equivalent classes; e.g., for /i, e/ it would prefer the "minimal" [−Back] over the "redundant" [−Back, −Low].
- **But there may be multiple minimal specifications for a given natural class.**

(10) Balearic Catalan vowel features (after Wheeler 2005:§2.2):

	High	Low	Front	Round	Tense
i	+	−	+	−	+
e	−	−	+	−	+
ɛ	−	−	+	−	−
a	−	+	−	−	−
ɘ	−	+	−	−	−
ɔ	−	−	−	+	−
o	−	−	−	+	+
u	+	−	−	+	+

(11) Minimal but non-unique intensional specifications:

a.	/i, e/:	[+Front, +Tense]	or	[−Round, +Tense]
b.	/u/:	[+High, −Front]	or	[+High, +Round]

- In contrast and as we demonstrate below, there is **exactly one maximal intensional specification for any natural class**.

8 More on the subset principle

(This is adapted from Hale and Reiss 2003 and Hale and Reiss 2008: ch. 2.)

- Phonological learning must proceed conservatively, since our assumptions (§1) make it impossible to "backtrack" from an overly general grammar.
- Maximal specification yields the minimal extension consistent with a natural class. For example, if one has only seen /m/ delete, one cannot yet assume that all nasals delete.

(12)
a. Earlier stage ⊆ later stage
b. [+Nasal, +Labial, −Coronal, −Dorsal] ⊆ [+Nasal]
c. /m/ ⊆ /m, n, ɲ, ŋ, .../

- Generalization occurs when:
  - additional positive evidence causes the learner to prune features from the intensional characterization, or
  - the structure of natural classes implies a segment not yet seen to undergo (or trigger) a rule must do so (see §A.2 for an example).
- Learners are *epistemically bounded* (in the sense of Fodor 1980:333f.); for example, if S and T are seen to trigger some rule R, then:
  - the learner must infer that the structural environment for R is $S \cap T$, and
  - if there is also some segment U such that $U \supseteq S \cap T$, the learner must infer that R's structural environment also includes U.

9 Feature maximization

- Let F be the universal, finite feature specification.
- A feature specification is a set of pairs (α, f) such that $\alpha \in \{+, -\}$ and $f \in F$.
- Let S and T be feature specifications for two segments.
- The maximal feature specification for $\{S, T\}$ is given by the intersection $S \cap T$.[4]

(13) **Algorithm:** intersection of feature specifications $R = S \cap T$:

$R \leftarrow \emptyset$
for $(\alpha, f) \in S$ **do**
    if $(\alpha, f) \in T$ **then**
        $R \leftarrow R \cup \{(\alpha, f)\}$
    end if
end for

[4]The resulting natural class may contain segments other than S and T; see §A.2 below for a scenario where this is a desirable outcome. Furthermore, if intersection produces $\emptyset$, this implies the two segments belong to the natural class which includes all segments, since $\emptyset$ is a subset of all other sets.

- This algorithm runs in linear time, specifically $O(|F|)$.
- This algorithm produces a unique solution: there is only one intersection of two sets.
- Since intersection is associative (and commutative), this algorithm can be generalized for more than two segments. That is, for feature specifications S, T, and U:

$$S \cap T \cap U = (S \cap T) \cap U = S \cap (T \cap U).$$

- Applying this algorithm to just the three contrastive vowel features in Georgian, as in (9), one obtains [−Back, −Low].
- Applying this algorithm to {i, e} using the 24-feature "universal" feature specifications provided by Hayes (2009:§4.10), one obtains a natural class with 23 features specified.

(14) $$\begin{bmatrix} -\textsc{Back} \\ -\textsc{Low} \\ -\textsc{Round} \\ -\textsc{Nasal} \\ \dots \end{bmatrix} = \left\{ \varsigma : \varsigma \supseteq \begin{Bmatrix} -\textsc{Back} \\ -\textsc{Low} \\ -\textsc{Round} \\ -\textsc{Nasal} \\ \dots \end{Bmatrix} \right\}$$

- This specification characterizes the environment of the Georgian lateral-fronting rule.[5]

10 The bottom line

The received wisdom that feature specifications are minimal faces serious problems. **Feature maximization is free from these problems.**

11 Georgian, again

- Arguably, the informal preference for minimal specification leads phonologists astray!
- Above we followed Robins and Waterson (henceforth, R&W) in their assumption that Georgian /a/—Mxedruli ⟨ა⟩—is +Back.
- However, R&W later describe this phoneme as being "generally of front quality" (*loc. cit.*) except before /u, ɫC, ɫ#/ or after /q, ɫ/, where they give it as [ɑ].
- *Pace* R&W, we assume their /a/ is underlyingly −Back, with a +Back allophone [ɑ].
- Thus, the "minimal" (7a) is not an empirically adequate environment for /ɫ/-fronting (barring further assumptions, e.g., about the orderings among the rules of allophony): a more specified structural environment—as in (7bc)—is necessary.

[5]Naturally, it contains specifications for all features except High.

12 Malayalam and assimilation bias

- This bias for minimality is an instance of—and interacts with—a general bias that common phonological rules should be the easiest to express.

> The goal of phonology is the construction of a theory in which cross-linguistically common and well-established processes emerge from very simple combinations of the descriptive parameters of the model. (McCarthy 1988:84)

- For instance, many phonologists suppose that special mechanisms—agreement, copy, linking, spreading, and so on—are needed to express assimilation and/or harmony (e.g„ Hyman and Schuh 1974, Hayes 1986, Rose and Walker 2004, Hansson 2007).
- In contrast, we assume that notions like *assimilation* are mere “taxonomic artifacts”:

> There are no rules for forming relative clauses in Hindi, verb phrases in Swahili, passives in Japanese, and so on. The familiar grammatical constructions are taken to be taxonomic artifacts, useful for informal description perhaps but with no theoretical standing. They have something like the status of “terrestrial mammal” or “household pet”. (Chomsky 2000:8)

- An important case is found in Malayalam. In this language simple velar onsets palatalize after /i, e, a/ but not after /u, o/.

(15) Malayalam velars:

a.	kutːi**k**ʲə	‘child (dat.)’	(Mandal in press)
	we**k**ːʲal	‘cooking’	
	kana**k**ʲam	‘gold’	
b.	en̪t̪u**k**oɳʈaːɳə	what.with.it	(Mohanan 1996:425)
	ɦrɨd̪roː**g**at̪t̪inor̪u	heart.disease.Acc.one	

- Mohanan and Mohanan (1984:586, fn. 24) give the exact same feature composition for Malayalam vowels as we gave for Georgian in (9), and formulate the palatalization rule as copying of −Back.

(16) Malayalam palatalization (after Mohanan and Mohanan, *loc. cit.*):

$$\begin{bmatrix} -\textsc{High} \\ -\textsc{Continuant} \end{bmatrix} \rightarrow \{-\textsc{Back}\} \;/\; \begin{bmatrix} -\textbf{Back} \\ +\textsc{Vocalic} \end{bmatrix} \text{—}$$

- But Mohanan and Mohanan add:

> Phonetically *a* is in fact a back vowel… (*loc. cit.*)

- *Pace* Mohanan and Mohanan, we assume that their /a/ is underlyingly +Back, and that the structural environment in (16) must be modified to conform to phonetic reality.

(17) Malayalam palatalization (revised):

$$\begin{bmatrix}-\textsc{High}\\ -\textsc{Continuant}\end{bmatrix} \rightarrow \{-\textsc{Back}\} \;/\; \begin{bmatrix}-\textbf{\textsc{Round}}\\ +\textsc{Vocalic}\end{bmatrix} \_\_$$

- This is no longer a feature-copying rule, but it is empirically adequate.

References

Chandlee, Jane, Rémi Eyraud, and Jeffrey Heinz. 2014. Learning strictly local subsequential functions. *Transactions of the Association for Computational Linguistics* 2:491–503.

Chen, Hubie, and Mans Hulden. 2018. The computational complexity of distinctive feature minimization in phonology. In *Proceedings of NAACL-HLT 2018*, 542–547.

Chomsky, Noam. 2000. *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Chomsky, Noam, and Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of Linguistics* 1:97–138.

Chomsky, Noam, and Morris Halle. 1968. *Sound Pattern of English*. Harper & Row.

Dell, François. 1980. *Generative Phonology and French Phonology*. Cambridge University Press.

Eisner, Jason. 1997. Efficient generation in primitive Optimality Theory. In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, 313–320.

Fodor, Jerry. 1980. Reply to Putnam. In *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*, ed. Massimo Piattelli-Palmarini, 325–334. Harvard University Press.

Frixione, Marcello. 2001. Tractable competence. *Minds and Machines* 11:379–397.

Garey, Michael R., and David S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company.

Gorman, Kyle. 2013. Generative phonotactics. Doctoral dissertation, University of Pennsylvania.

Hale, Mark, and Charles Reiss. 2003. The subset principle in phonology: why the tabula can't be rasa. *Journal of Linguistics* 39:219–244.

Hale, Mark, and Charles Reiss. 2008. *The Phonological Enterprise*. Oxford University Press.

Halle, Morris. 1978. Knowledge unlearned and untaught: what speakers know about the sounds of their language. In *Linguistic Theory and Psychological Reality*, ed. Morris Halle, Joan Bresnan, and George A. Miller, 294–303. MIT Press.

Hansson, Gunnar Ólafur. 2007. Blocking effects in agreement by correspondence. *Linguistic Inquiry* 38:395–409.

Hayes, Bruce. 1986. Assimilation as spreading in Toba Batak. *Linguistic Inquiry* 17:467–499.

Hayes, Bruce. 2009. *Introductory Phonology*. John Wiley & Sons.

Heinz, Jeffrey. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41:623–661.

Heinz, Jeffrey, Gregory M. Kobele, and Jason Riggle. 2009. Evaluating the complexity of Optimality Theory. *Linguistic Inquiry* 40:277–288.

Hyman, Larry M., and Russell G. Schuh. 1974. Universals of tone rules: evidence from West Africa. *Linguistic Inquiry* 5:81–115.

Idsardi, William. 2006. A simple proof that Optimality Theory is computationally intractable. *Linguistic Inquiry* 37:271–275.

Karp, Richard M. 1972. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, ed. Raymond E. Miller and James W. Thatcher, 85–103. Plenum.

Kenstowicz, Michael, and Charles Kisseberth. 1979. *Generative Phonology: Theory and Method*. Academic Press.

Linzen, Tal, and Gillian Gallagher. 2017. Rapid generalization in phonotactic learning. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8:1–32.

Mandal, Sayantan. In press. The cognitive phonetics of velar palatalization. In *Proceedings of WECOL 2022*, to appear.

McCarthy, John J. 1988. Feature geometry and dependency: a review. *Phonetica* 45:84–108.

Mohanan, K. P. 1996. Malayalam writing. In *The World's Writing Systems*, ed. Peter T. Daniels and William Bright, 420–425. Oxford University Press.

Mohanan, K. P., and Tara Mohanan. 1984. Lexical phonology of the consonant system in Malayalam. *Linguistic Inquiry* 15:575–602.

van Rooij, Iris. 2008. The tractable cognition thesis. *Cognitive Science* 32:939–984.

Odden, David. 2013. *Introducing Phonology*. Cambridge University Press, 2nd edition.

Rasin, Ezer, Iddo Berger, Nur Lan, Shefi, and Roni Katzir. 2021. Approaching explanatory adequacy in phonology using Minimum Description Length. *Journal of Language Modelling* 9:17–66.

Reiss, Charles. 2017. Substance free phonology. In *The Routledge Handbook of Phonological Theory*, ed. S.J. Hannahs and Anna Bosch, 425–452. Routledge.

Reiss, Charles, and Veno Volenec. 2022. Conquer primal fear: phonological features are innate and substance free. *Canadian Journal of Linguistics* 67:581–610.

Robins, R. H., and Natalie Waterson. 1952. Notes on the phonetics of the Georgian word. *Bulletin of the School of Oriental and African Studies* 141:55–72.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80:475–531.

Spencer, Andrew. 1996. *Phonology: Theory and Description.* Blackwell.

Volenec, Veno, and Charles Reiss. 2020. Formal generative phonology. *Radical: A Journal of Phonology* 2:1–148.

Wheeler, Max W. 2005. *The Phonology of Catalan.* Oxford University Press.

Zsiga, Elizabeth C. 2012. *The Sounds of Language: An Introduction to Phonetics and Phonology*. John Wiley & Sons.

A Responses to reviewers

A.1 Greed isn't good

- An anonymous reviewer argues Chen and Hulden's heuristics may represent viable algorithms for feature specification:

  > ...the branch & bound algorithm of Chen and Hulden does find the minimal feature specification, albeit with a fairly large search space. It is only the greedy algorithm that fails, but even then the failure is one of positing 4 features in a case where 3 would have sufficed. That is actually encouraging! These are general purpose algorithms, and a specialized feature minimization algorithm is likely to yield better results, for example by prioritizing coarse features like consonantal over more fine-grained ones like strident. Features are not the same as the arbitrary sets in the set cover problem, some features are more important than others. [...] **I fully expect the problem to be highly tractable under these parameters**.

- We submit: **why bother? There is simply no good reason**—for the linguist, or the Georgian infant—**to search for a non-unique, heuristic solution.**
- *Contra* the reviewer, exploiting implicational relations among features—e.g., the fact that [+Strident] implies [+Consonantal]—does not itself make set cover feasible.

A.2 But does it *Bachs*?

- Two anonymous reviewers ask if maximal specification can be reconciled with the following test:

  > The test we shall use is one suggested to me some years ago by Lise Menn. It consists of asking English speakers to form the plural of a foreign word that ends with a sound that does not occur in English. A good example, Ms. Menn suggested, is the German name *Bach* as in *Johann Sebastian* ___, which ends in the sound symbolized by /x/. (Halle 1978:102)

- It appears that English speakers, attempting a hyperforeign pronunciation of *Bachs*, do in fact produce [baxs] rather than *[baxz] or *[baxɨz].
- The stem-final segments which select the /-s/ plural are /p, t, k, f, θ/.[6] Applying (13) yields a natural class which can be informally described as the set of voiceless obstruents:

(18) $$\begin{bmatrix} -\text{Voice} \\ +\text{Consonantal} \\ -\text{Nasal} \\ -\text{Sonorant} \\ -\text{Lateral} \\ \ldots \end{bmatrix}$$

[6] This set does not explicitly contain /s, z, ʃ, ʒ/, because voice assimilation is bled by epenthesis when stems end in these consonants (see Volenec and Reiss 2020:30f. for discussion).

- /x/ is clearly a member of this natural class; i.e., its features are a superset of (18). **Thus maximal feature specification passes the *Bachs* test**.

A.3 Erratic phonotactic schematics

- An anonymous reviewer asks:

> I wonder how the author(s) would reconcile this learning model with the evidence that both children and adults seem to aggressively generalize phonotactic restrictions from limited data (e.g. just [p]) to larger, unobserved natural classes (e.g. [p f b v]). See e.g. the discussion in Linzen and Gallagher (2017). If those results are credible, they seem much more consistent with learning minimal feature specifications for natural classes than learning maximal ones.

- Note that Linzen and Gallagher (henceforth L&G) study phonotactic learning, whereas our proposal concerns phonological rule learning. We have independently critiqued standard assumptions in phonotactic theory (e.g., Gorman 2013:§2, Reiss 2017:§6).

- We also note that L&G's subjects are adults briefly exposed to an artificial language. It is not at all clear what such a study contributes to our understanding of child acquisition of real languages *in situ*.

- But let us grant, for sake of argument, that our proposal is also applicable to rapid artificial phonotactic learning in adults.

- Third, the reviewer is incorrect; the result from L&G they quote is not consistent with minimal specification. L&G hypothesize participants will construct "minimal classes":

> For example, when acquiring the phonotactics of English, learners may first learn that both [b] and [g] are valid onsets for English syllables before they can generalize to other voiced stops (e.g., [d]). This generalization will be restricted to the minimal class that contained the attested onsets (i.e., voiced stops), at least until a voiceless stop onset is encountered. (L&G:2)

- If by "minimal class" L&G refer to a natural class with the fewest segments, then presumably they would endorse our proposal, since **the smallest empirically adequate extension is given by a maximally specified intension**.

- It is unclear whether such a class would contain [d]. For instance, if major place features are bivalent, as we assumed above, then the intersection of the features associated with [b, g] will contain the specification [−Coronal] and exclude [d].[7]

- The matter is similarly unclear if we interpret "minimal class" intensionally, in terms of the number of features. **The minimal intensional specification for a single segment** (as in the reviewer's example) **will not generalize to any other phoneme**.

[7] L&G suggest that maximum entropy models may have this property, but provide no evidence.