1 Introduction

(1) Learning a language involves learning patterns which can be described in terms of logical relations between features. A couple of recent examples:

a. The Scottish Vowel Length Rule: \([i u ai]\) are long if and only if they precede a voiced fricative or \([r]\) or a morpheme boundary (Stuart-Smith, Thursday).

b. Northern Kammu tonotactics: If a syllable has a voiceless stop or a sonorant onset, then the syllable can bear either high or low tone; else if its onset is a glottal stop, then its tone is low; else its tone is high (Svantesson, Friday).

(2) This talk addresses an apparent mismatch between, on the one hand, the typological frequency of logical structures in natural-language phonology, and, on the other, the difficulty of learning analogous artificial patterns in the lab.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Logical structure</th>
<th>Lab</th>
<th>Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A \text{ if and only if } B) (regardless of (C))</td>
<td>Harder</td>
<td>More common</td>
</tr>
<tr>
<td></td>
<td>At least two of (A, B, C)</td>
<td>Easier</td>
<td>Less common</td>
</tr>
</tbody>
</table>

(3) Proposed solution: The typological frequency of different structures is affected by the abstract structure of the phonetic precursors.

(4) Outline of the talk:

§2 Lab studies of phonological pattern learning. SHJ Type II easier than Types III–V.

§3 Typology: Type II is common, Type IV rare. How come?

§4 Phonetic precursors of the SHJ types.

§5 Illustration: English Diphthong Raising.

§6 Discussion.

---

1Much of the work reported here is part of a continuing collaboration with Joe Pater and Katya Pertsova. I am indebted to several other colleagues for ideas, discussion, and critique, including Ricardo Bermúdez-Otero, Jeff Mielke, Jen Smith, and Erik Thomas, as well as to audiences at the Manchester Phonology Meeting and at NELS in 2012. Any remaining errors are to be ascribed to the author. Some of the work was funded by an internal grant from the University of North Carolina at Chapel Hill.
2 Structural bias in pattern learning

(5) Framework: What determines synchronic phonological typology?

a. Inductive bias (aka analytic bias), a property of the learner’s pattern-induction processes, makes some patterns easier to learn than others. Bias is inescapable for any inductive learner (see Pinker 1979; Mitchell 1990; Gallistel et al. 1991 for reviews. A similarity metric counts as an inductive bias; see Watanabé 1965; Quine 1969).

b. Channel bias, a property of the articulatory-acoustic-perceptual channel, systematically distorts the phonological form of utterance in transmission, thereby introducing new, phonetically-motivated patterns into the learner’s input (Hyman, 1976; Ohala, 1992, 1993). Bias is probably also inescapable in any real-world communication channel.

c. Together, inductive and channel bias skew language change, and hence also skew the long-term steady-state frequencies of different kinds of patterns (Bell, 1970, 1971; Greenberg, 1978).

Research problem: Infer humans’ inductive and channel biases by observing typology, lab experiments, acquisition, change, etc.

(6) Recent examples of proposals about inductive and channel biases:

a. Inductive bias: The phonological learner detects patterns based on sound environment, not on lexical category (Bybee, Thursday).

b. Channel bias: Individuals differ in the degree to which they coarticulate and compensate for coarticulation (Yu, Friday).

2.1 Kinds of bias

(7) Structure vs. substance: Here are three patterns which have the same structure but different substance:

<table>
<thead>
<tr>
<th>a. Phonology</th>
<th>b. Morphology</th>
<th>c. Non-linguistic game</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel</td>
<td>Number</td>
<td>Shapes</td>
</tr>
<tr>
<td>short</td>
<td>Case</td>
<td>Colors</td>
</tr>
<tr>
<td>long</td>
<td>sing.</td>
<td>One</td>
</tr>
<tr>
<td>*lám</td>
<td>mur</td>
<td>Illegal</td>
</tr>
<tr>
<td>lám</td>
<td>mur-s</td>
<td>Legal</td>
</tr>
<tr>
<td>*lám:m</td>
<td>mur</td>
<td>Legal</td>
</tr>
</tbody>
</table>

Swedish: Either the vowel or the consonant of a closed stressed syllable is long, but not both (Löfstedt, 1992).

Old French: /-s/ is attached to an o-stem noun if it is nominative or plural, but not both (Luquiens, 1909, §289).

Qwirkle: In a row of tiles, either the colors or the shapes must differ, but not both (Ross, 2006, 2).

(8) Much theoretical and experimental work in phonology has been motivated by the hypothesis that learners have a substantive inductive bias which favors phonetically-motivated hypotheses over others (Stampe, 1973; Prince and Smolensky, 1993; Archangeli and Pulleyblank, 1994; Hayes, 1999; Steriade, 2001; Wilson, 2006). Examples of hypothesized substantive biases:

a. There is a natural (innate) phonological process of final-obstruent devoicing, but not one of final-obstruent voicing or final-sonorant devoicing (Donegan and Stampe, 2009).

b. CON includes ONSET and NOCODA, but not NOONSET or CODA (Prince and Smolensky, 1993)
c. $\text{Dep[p t k]} / \_V$ universally dominates $\text{Dep[?]} / \_V$ (Steriade, 2001)

d. The plasticity of a markedness constraint depends on the perceptual similarity of the sounds in the largest change which the constraint can motivate (Wilson, 2006).

(N.B. I’m citing these hypotheses, not necessarily endorsing them.)

(9) Substantive inductive bias has loomed very large in linguistic theory, for at least two reasons:

a. One of the most striking trends in phonological typology is, it tends to look like phonetics. Substantive bias offers an explanation for why.

b. Since substantive bias applies to inductive problems that only arise in phonology, finding any substantive bias at all would be good evidence that there are cognitive specializations for phonology.

(10) But structural bias also has a history in linguistic theory. Examples of hypothesized structural biases:

a. A phonological rule can only add or delete an association line on a Feature-Geometric tier (McCarthy, 1988).

b. When two markedness constraints are equally effective at distinguishing licit from illicit forms, rank the one that refers to more features, or that uses more disjunctions, lower (slightly generalized from Gordon 2004).

c. Every learnable phonotactic pattern can be represented as a finite-state machine (Heinz and Idsardi, 2011).

(11) Why would linguists care about structural inductive bias?

a. It is informative about the architecture of the learner, e.g., rule-based vs. prototype-based learning (Medin and Schwanenflugel, 1981). (Sóskuthy, Thursday.)

b. The lab evidence for it is strong (see review in Moreton and Pater 2012a,b).

c. Structural bias is a potential link between phonology and other domains (morphology, syntax, psychology, biology, computer science, philosophy, . . . ). (Babel, McGuire, & Russell, Thursday.)

d. And . . .

(12) Most relevantly for this week’s theme, structural inductive bias might guide sound change

a. At the point of phonologization (this talk), or

b. In maintenance and transmission of a phonologized pattern over time (Seinhorst, today?)
2.2 Structural inductive bias in phonotactic patterns (Moreton et al., 2013, Exp. 1)

(13) Research in visual pattern learning has focused on a family of pattern types first studied by Shepard et al. (1961):

- **Type I** is a simple one-feature affirmation
- **Type II** is IFF/XOR on two features
- **Types III-V** need all three features, though subsets can be described with two
- **Type VI** is a three-way IFF/XOR; every subset needs three features

(N.B. We’re focusing on the SHJ types because there’s a big literature on them. The complete family of ways to divide a 2x2x2 stimulus space has been studied by Feldman (2000). There is also an extensive literature on 2x2 spaces, e.g., Bruner et al. (1956); Neisser and Weene (1962); Haygood and Bourne (1965). The SHJ patterns were first discussed in phonology by Silverman 1999.)

(14) Supervised learning (participant sees a shape, classifies it as A or B, receives right/wrong feedback, then on to the next one); no test of generalization outside the training set. The rate of learning decreased with the number of critical features: I > II > III, IV, V > VI. This result has been replicated many times in the vision literature (reviewed in Kurtz et al. 2013).

⇒ Not all of these patterns are equally easy to learn. The human learner’s inductive biases make some of them harder than others.

(15) Some recent visual studies have also found that the advantage for Type II over Type IV can be reduced or even reversed by changing the stimuli or task conditions:

- Using unsupervised training (Love, 2002)
- Not instructing participants to look for a rule (Love, 2002; Love and Markman, 2003; Lewandowsky, 2011; Kurtz et al., 2013)
- Making stimulus dimensions harder to verbalize, or perceptually less-separable (Nosofsky and Palmeri, 1996; Kurtz et al., 2013)

All of these changes make the task more like language learning.
(16) Stimulus design: MBROLA-synthesized $C_1V_1C_2V_2$ words defined by 8 binary features. (These stimuli have been used in several other experiments, including Moreton 2008; Lin 2009; Kapatsinski 2011; Moreton 2012)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Stimulus position $\sigma_1$</th>
<th>$\sigma_2$</th>
<th>Consonants</th>
<th>Vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced</td>
<td>$\pm$</td>
<td>$\pm$</td>
<td>$- - + +$</td>
<td></td>
</tr>
<tr>
<td>Coronal</td>
<td>$\pm$</td>
<td>$\pm$</td>
<td>$- + - +$</td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>$\pm$</td>
<td>$\pm$</td>
<td>$-$</td>
<td>$+ +$</td>
</tr>
<tr>
<td>back</td>
<td>$\pm$</td>
<td>$\pm$</td>
<td>$- + - +$</td>
<td></td>
</tr>
</tbody>
</table>

(17) For each participant, 3 of the 8 stimulus features were randomly chosen as the relevant features, and then randomly mapped onto the three logical features defining the Shepard pattern to produce the “language” for that participant. Examples of artificial phonotactic patterns instantiating SHJ Types I, II, and IV.

<table>
<thead>
<tr>
<th>L1 (TYPE I): C1 is voiced</th>
<th>digu, gada, dika, gugu, ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 (TYPE II): C1 is voiced iff V2 is back.</td>
<td>digu, tagi, kagaæ gada, ...</td>
</tr>
<tr>
<td>L3 (TYPE IV): At least two of: C1 is voiced, V2 is high, V2 is back</td>
<td>kaku, digu, guki, dæka, ...</td>
</tr>
</tbody>
</table>

(18) N.B. This procedure deliberately generated “crazy rules” (Bach and Harms, 1972; Anderson, 1981), phonetically unmotivated and typologically unattested, since the goal of the experiment was to study purely structural effects. ( Phonetic motivation and typological frequency in any case have weaker effects than those of structure; Moreton and Pater 2012a,b.)

(19) Procedure:

a. **Instructions:** Participants (who were run in a lab, by a human) were told they would learn to pronounce words in an artificial language, and then be tested on ability to recognize words in that language.

b. **Familiarization:** They listened to and repeated aloud 32 randomly-chosen pattern-conforming stimuli 4 times over.

c. **Test:** Then they heard 32 randomly-chosen pairs of new stimuli (one pattern-conforming, one not) and tried to identify the one that was “a word in the language you were studying”.
This figure shows the proportion of pattern-conforming responses for each participant (Moreton et al., 2013, Figure 8).\(^2\)

Analysis by mixed-effects logistic regression (\texttt{lmer} in R), with Type II as reference category and a random intercept for Participant.\(^3\) Fixed-effects part of fitted model (4608 responses from 144 participants; log-likelihood \(= -2879\)):

| Coefficient          | Estimate | SE   | z    | Pr(>|z|) |
|----------------------|----------|------|------|----------|
| (Intercept)          | 0.12803  | 0.14337 | 0.893 | 0.371865 |
| I                    | 0.81999  | 0.17583 | 4.663 | < 0.0001 |
| III                  | 0.47202  | 0.17322 | 2.725 | 0.006432 |
| IV                   | 0.63399  | 0.17423 | 3.639 | 0.000274 |
| V                    | 0.38984  | 0.17265 | 2.258 | 0.023952 |
| VI                   | -0.06134 | 0.17119 | -0.358 | 0.720114 |
| Both Same Genus      | 0.11078  | 0.24329 | 0.455 | 0.648850 |
| Both Same Segment    | 0.69318  | 0.25149 | 2.756 | 0.005845 |
| CorrFirst            | 0.27396  | 0.06348 | 4.316 | < 0.0001 |
| Redup                | -0.77231 | 0.10142 | -7.615 | < 0.0001 |

Unlike SHJ, Type II is harder than Types III–V.

\(^2\)Plotting symbols for Type II: + = all relevant features are in the same segment analogue; \(\triangle\) = all relevant features are of the same type (agreement/disagreement pattern); \(\circ\) = other.

\(^3\)Since repeated measures were not made on items, no random intercept was included for Item.
3 The typological conundrum

(23) Problem: This result doesn’t seem to square with natural-language typology. In the next thirty seconds,

a. If you were born in an *odd-numbered year*, please write down as many natural-language phonotactic patterns as you can think of that have the form “*A if and only if B*”, or “*A unless B*”.

b. If you were born in an *even-numbered year*, please write down as many as you can that have the form “*At least two of A, B, C*”, or “*At most two of A, B, C*”.

(24) Some of this (apparent) asymmetry may be due to a special subclass of Type II patterns, the agreement and disagreement patterns, which are both more frequent in nature, and easier to learn in the lab, than other Type II patterns (Moreton, 2008; Lin, 2009; Moreton, 2012). However, that can’t be the whole story:

a. There are still many iff/xor distributional patterns involving instances of two different features (typical Ling 101 complementary distribution patterns)

b. This residue is especially noteworthy since feature systems are deliberately engineered to make iff/xor patterns look like agreement (the idea behind Feature Geometry).

c. If you take the phonologically active classes in P-Base (Mielke, 2008), coerce them to SHJ types relative to the SPE feature system, and compare the results to a chance model, Type II is more overrepresented than Types III–V (Moreton and Pertsova, 2012). This can’t possibly be due to a bias for agreement/disagreement patterns, since all the relevant features co-occur in a single segment.

(25) ⇒ There is apparently a mismatch between the relative difficulty of Type II and Types III–V in the lab on the one hand, and their relative frequency in natural-language typology on the other hand. Possible responses:

a. “Pattern learning in the lab has nothing to do with natural-language typology.”

b. “The typological asymmetry in favor of Type II isn’t coming from inductive bias, so it must be coming from channel bias”, i.e., from the phonetic precursors.

Both are worth considering, but today’s talk will focus on the second one.

4 Structural bias in phonetic precursors

(26) If the (apparent) typological bias isn’t coming from inductive bias, where is it coming from? Maybe it’s inherited from the phonetic precursors: If phonetic variables tend to interact with each other in ways that are analogous to Type II rather than to Types III–V, then the available pool of phonetic variation could be skewed enough towards Type II to overcome inductive bias in the other direction.
Scenario: Formation of three-feature distributional patterns starting from

a. Two already-phonologized phonetic variables $x$ and $y$, instantiating phonological features $X$ and $Y$ with means $x_+, x_-$ and $y_+, y_-$ and standard deviations $\sigma_x, \sigma_y$. (For simplicity, assume no covariance between $x$ and $y$.)

b. One unphonologized phonetic variable $z$

Assumption: Phonologization consists of

a. Introducing a binary phonological feature $Z$ that divides the phonetic $z$ axis to create a $2\times2\times2$ phonological partition of the $(x, y, z)$ space

b. Designating densely-populated cells as “permitted” and sparsely-populated ones as “forbidden”.

Put together two ideas:

a. Minimum-difficulty surface (Lindblom, 1990): For a given $(x, y)$, $E[z]$ defaults to the corresponding point $f(x, y)$.


Example: Phonologization to Type II. Note cartoon projections of the marginal $z$ distribution at left. The equatorial plane (added in phonologization) falls on a notch in the distribution.
(31) This figure illustrates an important special case, namely, the one where the minimum-difficulty surface is a (possibly tilted) plane. I.e., \( f(x, y) \) is a linear function of \( x \) and \( y \).

We might expect \( f \) to often be linear, or nearly so, since real-world functions are differentiable, so a small enough piece is close to being a (possibly tilted) plane.

(32) A linear \( f \) restricts the possible ways to get a bi- or multimodal marginal distribution of \( z \). The possibilities are:

a. A bimodal distribution in which two adjacent \( x-y \) cells cluster together with a high \( z \) and the other two with a low \( z \), as in (30)

b. A unimodal, or broadly smeared, distribution in which there is no clear separation. Phonologization would create a Type I pattern.
c. A trimodal distribution in which the ++ and -- cells cluster in the middle, with ++ above and -- below. Phonologization would create a Type V pattern.

(33) The $(X,Y)$ categories are four points on a plane. Phonologization draws a straight line across that plane, cutting off one point or two adjacent points. This holds even if, e.g., the $x$ and $y$ axes aren’t parallel.
Not to say that Type IV would have no possible precursor in this scenario; it would just have to be very non-linear:

(with the \((-X, -Y)\) combination ruled out by some independent factor).

In \(XYz\) phonologization scenarios, then, the precursor pool is predicted to be skewed towards Types I, II, and V, perhaps enough so to overcome inductive bias against Type II.

### 5 Example: The English Diphthong Raising family

For a real-life illustration of these patterns, we turn to a family of alternations in English around the world, in which the historical /\(aI\)/ diphthong has higher vs. lower allophones depending on following context (TIGHT/TIDE/PRICE/PRIZE):

|        | SVLR \(-\text{cont}\) & +\text{cont} | SGW \(-\text{cont}\) & +\text{cont} | CR \(-\text{cont}\) & +\text{cont} |
|--------|----------------------|--------------|----------------------|--------------|--------------|
| −voice | \(\ddot{a}i\)         | \(\ddot{a}i\) | \(ai\)               | \(ai\)       | \(\ddot{ai}\)  | \(\ddot{ai}\) |
| +voice | \(\ddot{ai}\)         | \(ai\)       | \(a:\)               | \(a:\)       | \(ai\)         | \(ai\)       |

Good example of phonologization because

a. Versions of it keep getting independently re-innovated, even in the 20th Century (see review in Moreton and Thomas 2007).

b. Has lexical exceptions (Vance, 1987)

c. The difference between the diphthongs is contrastive before flap (Mielke et al., 2003; Boersma and Pater, 2007; Pater, 2014)

Phonetic basis of CR/SGW: Abbreviation of diphthong nuclei and exaggeration of offglides before voiceless codas (Thomas, 2000; Moreton, 2004).
(39) What about the SVLR-related height alternation? Bermúdez-Otero (2014) suggests uniting it with CR/SGW via Pre-Fortis Clipping. Here is an alternative hypothesis: Diphthongs are also more open (lower) before fricatives than before stops. Test this with [er] in West Coast U.S. English to maximize chances of seeing unphonologized phonetic effect.

(40) Isolated-word production study of 16 American English speakers, mostly from California (exceptions: 1 moved around a lot, 1 Ohio, 1 NJ, 1 Oregon). Read word list aloud in 3 different random orders:

<table>
<thead>
<tr>
<th>Context</th>
<th>Target vocoid</th>
</tr>
</thead>
<tbody>
<tr>
<td>[cont]</td>
<td>[voiced]</td>
</tr>
<tr>
<td>− −</td>
<td>− t</td>
</tr>
<tr>
<td>− +</td>
<td>− d</td>
</tr>
<tr>
<td>+ −</td>
<td>− s</td>
</tr>
<tr>
<td>+ +</td>
<td>− z</td>
</tr>
</tbody>
</table>

Dependent measure was $F_1$ at the $F_1$ maximum, which normally occurred early in the vowel. Measurement made in Hz with Praat (Boersma and Weenink, 2010), converted to Bark.

(41) Quadratic surface fit to the four condition means (R Development Core Team, 2005):

(42) Linear mixed-effects model (lmer function in lme4 package; random effects (1 + Continuant* Voice | Participant))

| Coefficient      | Estimate | SE  | $\chi^2$ | Pr(>|$\chi^2$|) |
|------------------|----------|-----|----------|-----------------|
| (intercept)      | 4.62677  | 0.11029 |          |                 |
| Continuant       | 0.12757  | 0.04064 | 9.0658   | <0.003          |
| Voice            | 0.26426  | 0.04811 | 45.3538  | <0.001          |
| Continuant:Voice | -0.07348 | 0.04394 | 2.2180   | 0.1364          |
Results jibe with expectations about precursor:

a. Voicing is associated with higher $F_1$ (lower articulations), as in Canadian Raising, Southern Glide Weakening, and similar patterns.

b. Continuancy is also associated with higher $F_1$, as in the Scottish [ai] pattern.

c. The precursor surface is approximately planar, and is oriented so that there is a vertical gap between the voiced and the voiceless obstruents, as in the (very common) CR/SGW, not the (apparently unique) Scottish pattern.

6 Discussion

Summary:

a. Inductive bias: Human phonotactic pattern learners are worse at learning Type II (iff/xor) patterns than at Type IV (at least two of . . . )

b. Typology: Although quantitative data is scarce, it does seem that Type II patterns are more common cross-linguistically than Type IV patterns. Hence, typology mismatches inductive bias.

c. Channel bias: In the $XYz$ phonologization scenario, if minimal-difficulty surfaces are approximately planar, then we expect phonologization to produce Type I, Type II, or Type V (depending on the disposition of the $XY$ points and the orientation of the surface).

⇒ Phonetic precursors have structural properties, and those structural properties may explain typological facts that are not explained by structural inductive biases.

Some questions raised:

a. What do the structural effects on pattern-learning difficulty reveal about the architecture of the learner? E.g., Type II vs. Type IV is often used as an indicator of rule-based vs. cue-based learning (Shepard et al., 1961; Gluck and Bower, 1988; Ashby et al., 1998; Love, 2002; Smith et al., 2012).

b. What are the logical structures of phonetic precursors, and how frequent are they? In the $XYz$ scenario, under what circumstances can we trust the minimum-difficulty surface to be approximately planar?

c. What, in fact, are the relative typological frequencies of different logical structures among phonological patterns?

d. Are the same logical-structure effects seen in phonologization and in the historical development of an existing pattern (e.g., generalization, (Cristiá et al., 2013))?
References


Donegan, P. J. and D. Stampe (2009). Hypotheses of Natural Phonology. MS, University of Hawai‘i.


Smith, J. D., M. E. Berg, R. G. Cook, M. S. Murphy, M. J. Crossley, J. Boomer, B. Spiering, M. J. Beran,


