Substantive analytic bias

Elliott Moreton
UNC-Chapel Hill

(1) Summary so far:

a. Asymmetries in phonological typology emerge from asymmetries in phonological change ...

b. ... which in turn emerge from biases in the production–perception–analysis loop:

(i) Channel bias: Systematic errors in the production–perception channel that introduce systematic differences between the phonological representations generated by other speakers and those received by the learner

(ii) Analytic bias: Systematic differences in learning response to data instantiating different patterns with equal statistical quality.

c. The existence of both kinds of bias is not in doubt; rather, it is their nature (what are they specifically?) and their typological effectiveness (what, if anything, do they contribute to typology?) that we have to worry about.

d. We've seen two kinds of channel bias so far:

(i) Paradigmatic simplicity bias:

i. It is easier to learn to distinguish two stimulus classes when the distinction is phonetically systematic ("featural") than when it is phonetically arbitrary.

ii. Evidence for existence is pretty strong (lots of studies, lots of different classes, tasks, native languages, ages, etc.)

iii. Evidence for typological effectiveness is weak (confounded with channel bias).

(ii) Syntagmatic simplicity bias:

i. It is easier to learn to distinguish two stimulus classes when the distinction depends on two instances of the same feature within the stimulus than when it depends on instances of two different features.

ii. Evidence for existence is decent, but scanty; only L1 English speakers, only adults.

iii. Evidence for typological effectiveness is better than for paradigmatic s.b., but limited to a single case, and there are objections to that.

(2) Both biases we've seen evidence for so far are *formal*, in the sense that they do not care about the real-world interpretation of the features.

a. Paradigmatic s.b.: Any feature can be the basis of a class.

Is this true? LaRiviere et al. (1974) only found featural/arbitrary differences for [strid] and [nas], but not for [cont] or [voice]. Does that indicate an analytic bias favoring the former two contrasts over the latter two?

b. Syntagmatic s.b.: Any feature (or pair of features) can define a within-stimulus dependency.

Evidence: Height-height and voice-voice act alike, as do height-backness and place-voice (data from last time).

(3) Does the real-world phonetic interpretation of a phonological pattern make a difference in how it is acquired? Both answers are live options.

(4) YES:

> The proposal is to let a distinct grammatical component, which I call the P-map (P for perceptibility) project correspondence constraints and determine their ranking. The P-map is a set of statements about relative perceptibility of different contrasts, across the different contexts where they might occur. For instance, the P-map will be the repository of the speaker's knowledge that the [p]-[b] contrast is better perceived before V's (e.g. in [apa] vs. [aba]) than before C's (e.g. in [apta] vs. [abta]).
>
> (Steriade, 2008, 151)

(5) NO:

> . . . [T]he substance of phonological entities is *never* relevant to how they are treated by the phonological system, except in *arbitrary, stipulative* ways. . . . [M]any of the so-called *phonological universals*. . . are in fact epiphenomena deriving from the interaction of extragrammatical factors like acoustic salience and the nature of language change. . . . Phonology is not and should not be grounded in phonetics since the facts that phonetic grounding is meant to explain can be derived without reference to *phonology*.
>
> (Hale and Reiss, 2000, 162)

(6) The existence of substantive analytic bias is an important question, since it bears on

a. the domain-specificity of phonological acquisition (is there something involved in learning phonology that isn't involved in learning anything else?)

b. the interpretation of typological data (if we see what appears to be an effect of channel bias, can we safely conclude that it is an effect of channel bias?)

(7) The most unambiguous example of a substantive analytic bias would be one between $S_1 \rightarrow S_2$ and $S_1 \rightarrow S_3$ in which $\mathbf{H}$ in effect did nothing but permute features.

Example: Final-obstruent voicing vs. final-obstruent devoicing (Steriade, 1997; Yu, 2004; Blevins, 2006b; Kiparsky, 2006; Blevins, 2006a). Simply invert the voicing feature of every final obstruent:

S_2	[kɔnzonant]	[voːkaːl]	[ʃtɪmbant]	[aʊslaʊt]
S_3	[kɔnzonand]	[voːkaːl]	[ʃtɪmband]	[aʊslaʊd]

1 Case study: Velar palatalization (Guion, 1998; Wilson, 2006)

(8) Velar palatalization is on everyone's list of common sound changes (Hock, 1991; Trask, 1996; Crowley, 1997; Campbell, 2004). The classic example is from English (Guion, 1996, Chapter 2):

Old English	Modern English	Gloss
kirike	tʃɹtʃ	'church'
kiːdan	tʃaɪd	'chide'
keːake	tʃik	'cheek'
ɣinian	jɔn	'yawn'
ɣeard	jɑɹd	'yard'
ɣeornan	jɹn	'yearn'

(9) Two implicational universals, true both synchronically and diachronically:

a. Velar palatalization before any vowel implies palatalization before all fronter vowels.

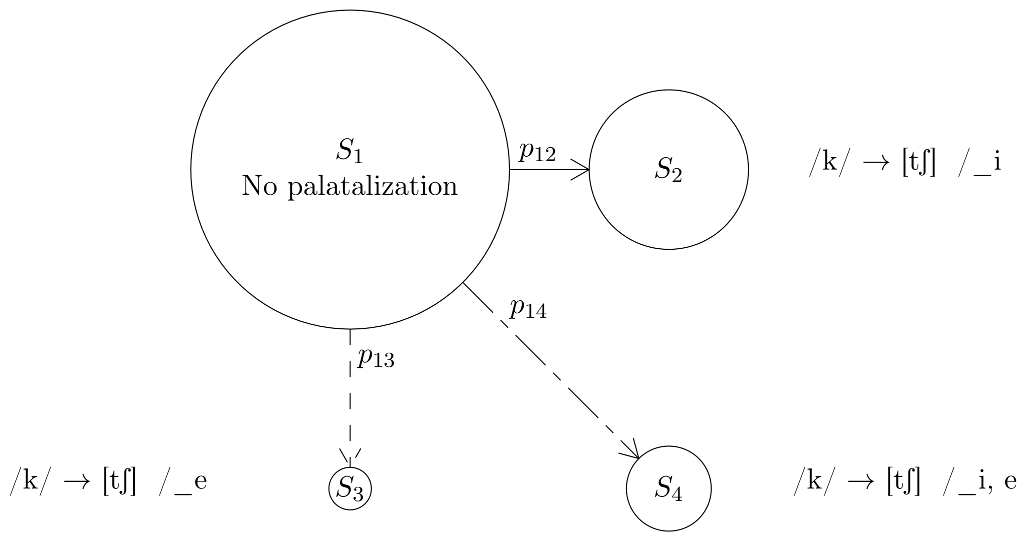
b. Palatalization of voiced velars implies palatalization of voiceless ones.

The opposite change, of tʃ→k or dʒ→g , is thought to be very rare (Guion, 1998; Chang et al., 2001). Here are some more palatalization examples, from Guion (1996, 1998):

Language	Change	Environment
Slavic (1st palatalization)	k→tʃ g→ʒ x→ʃ	j, ĭ, i, e, ɛ, ẽ
Slavic (2nd palatalization)	k→tʃ d→dʒ x→s	i, ɛ
Indo-Iranian	k→tʃ d→dʒ g^{h}→dʒ^{h}	i, e
Cowlitz Salish	k→tʃ k'→tʃ' x→ʃ	i
Bantu (I)	k→tʃ g→dʒ	j
Bantu (II)	k→ts g→z̤	i
English (Old→Middle)	k→tʃ ɣ→j kk→tʃ gg→dʒ	ææː e eː i ɪ
Mam (Mayan)	k→tʃ, tɹ̥, c	i, e
Chinese (Old→Middle)	k→tɕ k^{h}→tɕ^{h} g→dʑ x→ɕ	ji, je

1.1 Channel bias (Guion, 1996, 1998)

(10) The channel-bias proposal focuses on the rates of innovation of velar palatalization in different environments:



(11) Why does palatalization preferentially happen in these particular environments? Guion's proposed channel-bias answer, based on Ohala (1993):

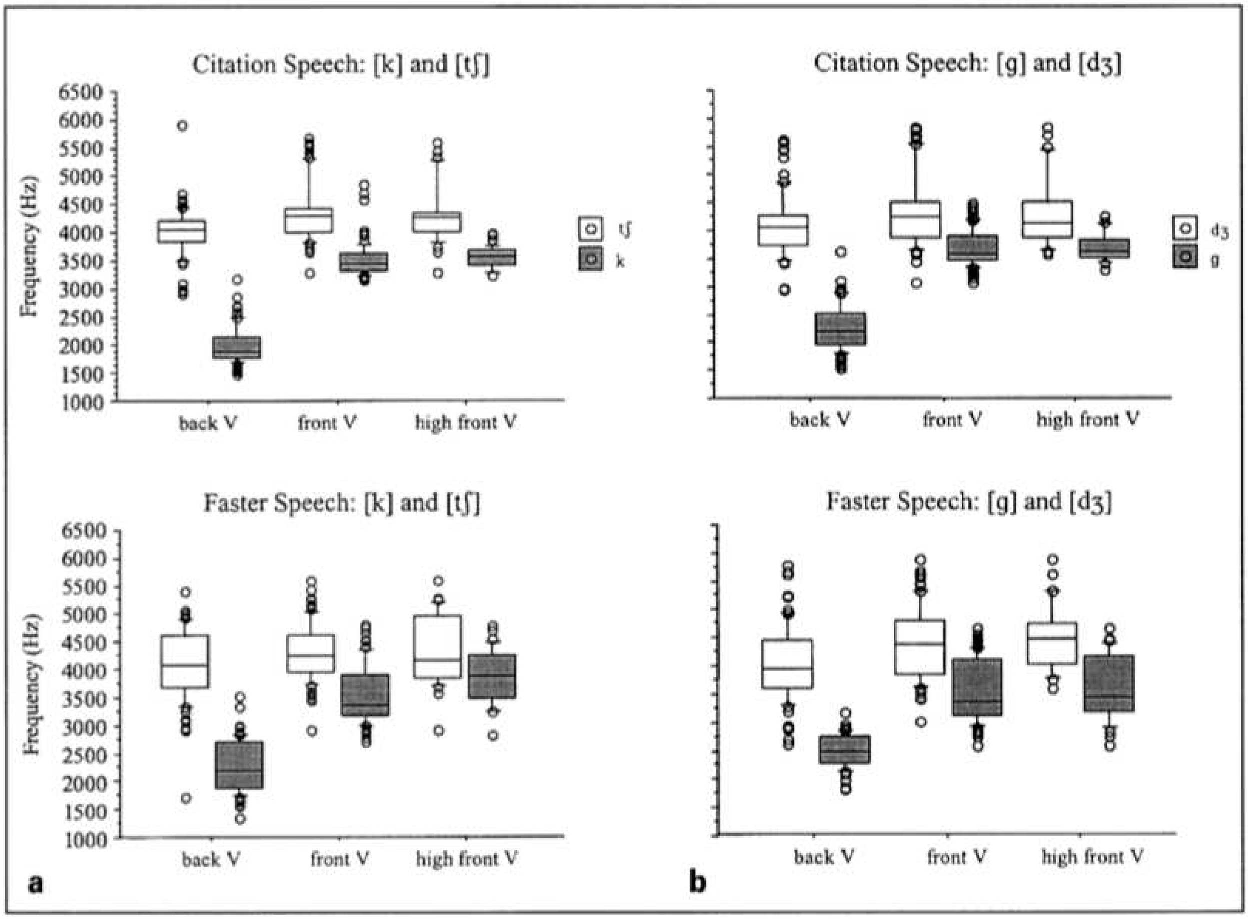
a. k→tʃ happens when the speaker says [k], but the listener parses it as [tʃ].

b. The environments that favor velar palatalization are the ones where [k] and [tʃ] are most easily confused.

(12) [k] and [tʃ] are distinguished by several acoustic properties, but let's just focus on one: The peak frequency of the release burst.

(Guion (1998) presents similar results for F_2, which I'll skip in the interests of time.)

(13) Results from 4 female American English speakers reading 42 English monosyllables with [k g tʃ dʒ] in the environments #_ [i ɪ eɪ ɛ æ ɑ ɔ oʊ ʊ u ʌ]. Measured the strongest frequency in the burst and aspiration part. Boxes enclose middle 50% of measurements; line in middle of box is median.



a. Velars and palato-alveolars are more similar before front, especially high front, vowels.

   Could be why (phonological) velar palatalization happens in front-vowel contexts rather than back-vowel ones, and why the trigger is especially likely to be a high front vowel.

b. Vowel context affects velars, not palato-alveolars.

   Could be why velars become palato-alveolars, rather than the other way around.

(14) Guion also did a perception experiment: She played *CV* syllables of the form [k tʃ g dʒ] + [i ɑ u] in background noise, and asked listeners to indicate which of [k tʃ g dʒ] they heard each time. Results:

Heard	Spoken											
	k			tʃ			g			dʒ		
	i	ɑ	u	i	ɑ	u	i	ɑ	u	i	ɑ	u
k	43			10			4			9		
		84			10			4			2	
			46			13			5			11
tʃ	35			85			4			28		
		13			87			–			23	
			31			84			3			22
g	10			–			71			12		
		3			–			87			10	
			12			–			76			17
dʒ	12			5			21			51		
		–			3			9			65	
			11			3			16			50

a. Velars are more likely to be heard as palatals before [i] than [a] . (But [u] unexpectedly behaves more like [i] than like [a] .)

b. [k] is more likely to be heard as [tʃ] than [g] is as [dʒ] .

Could be why (phonological) velar palatalization befalls /k/ more than /g/ .

1.2 Analytic bias (Wilson, 2006)

(15) Loose ends of the channel-bias account:

a. The acoustic and perceptual data from Guion's experiments provides only limited evidence for the crucial difference between [i] and [e] (only for voiceless segments in faster speech by female speakers in the production experiment).

b. The channel-bias account implicitly assumes that the learner responds alike to training data instantiating palatalization before [i] and [e] . But maybe palatalization is more noticeable before [e] (since velars and palato-alveolars differ more before [e]), leading learners to assign more weight to evidence for pre- [e] palatalization.

(16) Wilson suggests a mechanism of analytic bias:

a. Markedness constraints are restricted to specific contexts; faithfulness constraints are general.

b. Constraints are weighted rather than ranked.

c. Markedness constraints which motivate more-perceptible changes are more resistant to reweighting (i.e., require more learning data to move from the initial state).

d. Prediction: If the learning data supports a particular ranking, say, M_1 is 10 units heavier than F, and M_1 is very hard to move, then F will have to do most of the moving, with the result that it may end up below some other markedness constraint M_2.

⇒ Training someone to palatalize in a context where palatalization is very perceptible (such as before _e) may inadvertently cause them to learn to palatalize in a context where palatalization is not very perceptible (such as before _i).

(17) Experiment 1: Generalization on the context.

a. Relevant typological generalization: Palatalization before [e] implies palatalization before [i] .

b. Idea: Train English speakers on a language game with X. . . Y stimuli like [keɪnə. . . tʃeɪnə], and test on generalization to, e.g., [kina. . .].

c. Stimuli were CVCə nonwords, V ∈ {i eI ɑ}; C_1 ∈ {p b k g} for X, {p b tʃ dʒ}; irrelevant C_2 ∈ {p b k g m n f v θ ð s z tʃ dʒ l r w}.

d. 4 phases in each condition (prac, familiarization, break, testing). Exposure was "I say X, you say Y", with S repeating the Y. 32 trials. 2-min break with distractor task, then test like exposure but computer only says X. 80 trials. Same for both conditions. For trained contexts, half of test items were identical to some of the familiarization items.

e. 22 native Am Eng spkrs. Results transcribed, not by author.

	Training		Test, Pr(Pal)	
Stimulus	High	Mid	High	Mid
kiCə	tʃiCə		0.44	0.20
giCə	dʒiCə		0.52	0.48
keɪCə		tʃeɪCə	0.13	0.19
geɪCə		dʒeɪCə	0.14	0.49
kaCə	kaCə	kaCə	0.05	0.15
gaCə	gaCə	gaCə	0.14	0.39

f. The rate of pre- [i] palatalization significantly exceeded that of pre- [eɪ] palatalization in the High condition, but not in the Mid condition. W. interprets this as evidence of generalization from the pre- [eɪ] to the pre- [i] context, with no generalization in the other direction.

g. Alternative interpretations?

(18) Exp. 2: Generalization on the focus.

a. Relevant typological generalization: Palatalization of /g/ implies palatalization of /k/ .

b. The specific constraint set used here predicts that you shouldn't get generalization from [k] to [g] or vice versa. Exp. 2 tests that prediction.

c. Like Exp. 1, but with the stimuli re-paired. [Table corrected from original handout.]

	Training		Test, Pr(Pal)	
Stimulus	Voiceless	Voiced	Voiceless	Voiced
kiCə	tʃiCə *4	tʃiCə *1	0.39	0.26
keɪCə	tʃeɪCə *4	tʃeɪCə *1	0.36	0.20
giCə	dʒiCə *1	dʒiCə *4	0.14	0.50
geɪCə	dʒeɪCə *1	dʒeɪCə *4	0.11	0.44
kaCə	kaCə	kaCə	0.12	0.00
gaCə	gaCə	gaCə	0.09	0.23

d. As expected, those trained on voiceless didn't generalize to voiced, and vice versa.

(19) Why is palatalization more common in [g] than [k] typologically? W.'s answer: Channel bias (p. 972):

a. VP develops by perceptual reanalysis of coarticulation between front vowels and [k g] .

b. Coarticulated [k] sounds more like [tʃ] than coarticulated [g] does like [dʒ] (Guion, 1998).

c. For any given degree of coarticulation, [k] is more likely to undergo reanalysis than [g] .

1.3 Comments

(20) Could participants have been primed for asymmetric learning by English phonology?

a. People in both conditions would know whatever patterns L1 English had taught them; the difference would have to come from different *response* to the /tʃi/ and /tʃeɪ/ training data.

b. Maybe they had somehow learned that palato-alveolars are relatively scarcer before /eɪ/ than before /i/ , and had some predisposition to conclude from this that palatalization before /eɪ/ implies palatalization before /i/ .

c. Counts of syllable-initial *CV* from CELEX (Baayen et al., 1995)[1] are consistent to some extent:

	/j/	/i/	/ɪ/	/eɪ/	/ɛ/	/æ/	/a ɒ/
/tʃ/	0	8326	33219	16060	5160	7293	10417
/k/	34206	11524	81528	61752	6176	86080	126719
ln(tʃ/k)	$-\infty$	–0.33	–0.90	–1.35	–0.18	–2.50	–4.72
/dʒ/	80	6013	61930	6322	32332	4834	13538
/g/	8936	234	36464	19884	63657	16894	40025
ln(dʒ/g)	–4.72	3.25	0.53	–1.14	–0.68	–1.25	–1.08

d. /tʃ/ is relatively more likely before /i/ than /eɪ/ , and /dʒ/ likewise. The /i/ – /eɪ/ difference is greater for /k tʃ/ than for /g dʒ/ , which might account for the greater /eɪ/ -to- /i/ generalization in the voiced than the voiceless stimuli in Exp. 1 (see Wilson's footnote 13).

e. But then why wouldn't we see more [k] -to- [g] generalization in the [i] than the [eɪ] environment in Exp. 2? Training on /ki/ → [tʃi] should be strong evidence for /gi/ → [dʒi] .

(21) Re extension from e to ae: Something like this may have happened in Zuni; see Buckley (2000).

[1] From the file `EPW.CD`, which distinguishes different inflected forms. Counts were weighted by form frequency, but the pattern of results is substantially the same if unweighted counts are used.

2 Case study: Syllable structure (Schane et al., 1974)

(22) It is typologically common to delete C_1 from $VC_1{+}C_2V$ or $VC_1\#$ (Wilson, 2001), but probably much less common to delete it from $VC_1 + V$.

Maori (Kenstowicz and Kisseberth, 1979, 171)				French (Schane et al., 1974)		
Active	Passive	Gerundive				
wero	werohia	werohaŋa	'stab'	$\ldots VC + V \ldots$	$\ldots VC + C \ldots$	
hopu	hopukia	hopukaŋa	'catch'	petit ami	peti garsɔ̃	'little friend/boy'
aru	arumia	arumaŋa	'follow'	groz ɔ̃kl	gro pɛr	'big uncle/father'
mau	mauria	mauraŋa	'carry'	lɔ̃k ete	lɔ̃ prɛ̃tɑ̃	'long spring/summer'
afi	afitia	afitaŋa	'embrace'			

(23) Artificial-"language" task, based on French liaison rule:

a. 3 "adjectives", all ending in a consonant (/ˈtupɑk/ 'small', /ˈɑmuf/ 'white', /ˈgoʊmeɪt/ 'old')

b. 4 "nouns", beginning with either C or V (/ˈsipu/ 'man', /ˈpɑʃi/ 'house', /ˈoʊgɑ/ 'book')

c. Two conditions:

	$/C_1\#C_2/$ /ˈtupɑk ˈsipu/	$/C_1\#V_2/$ /ˈtupɑk ˈogɑ/
"Natural"	$\rightarrow [\#C_2]$ [ˈtupɑ ˈsipu]	[ˈtupɑk ˈogɑ]
"Unnatural"	[ˈtupɑk ˈsipu]	$\rightarrow [\#V_2]$ [ˈtupɑ ˈogɑ]

d. Neither rule is found in participants' native language (English); r-sandhi and *a/an* allomorphy "do[...] not permeate large parts of the English phonological system" (Schane et al., 1974, 353).

[Comment in class from Sverre Johnsen: English *does* have a tendency towards gestural overlap of C_1C_2 sequences when the Cs are stops, which could have primed acquisition of the "Natural" rule in the stop stimuli.]

(24) Methods:

a. 31 undergraduates, Midwestern AmEng, no r-sandhi. 24 with no French [these are the ones that the results below refer to unless otherwise specified], 7 with good French (4 yrs HS or more).

b. Ss instructed to learn and told they would be paid by learning rate. Error counter visible to S; red light flashes on error.

c. Stimuli produced using English phonetics; syllable boundary aligned to word boundary in two-word phrases. Audio only.

d. Training procedure:

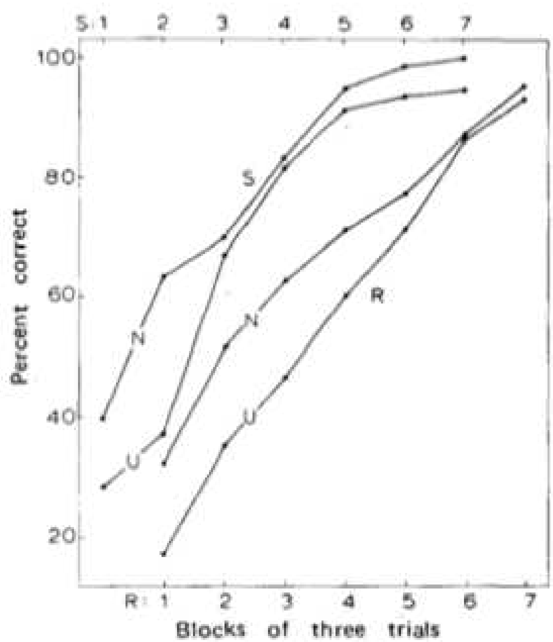
(i) Part A: Nouns in isolation. 10 random lists of the nouns for paired-associate learning: English prompt; 2-sec pause to produce translation, correct translation as model, participant repeats after model. N sets repeated over and over in different orders. Criterion was: two series with no errors.

(ii) Part B: A-N phrases. Ss were sorted into 4 groups to equalize performance in Part A: Natural Systematic, Natural Random, Unnatural Systematic, Unnatural Random.

All of the 12 possible A-N combinations were presented. R groups got random order of presentation; S groups got same adj with all 4 nouns before going on to next adj, and, in the first block, alternated between deleting and non-deleting trials. Same task as in A. Criterion was one block of 12 items with no errors. Any trial with "appropriate syllable structure" was scored as correct, even if adj was partly or wholly wrong, as long as noun was correct.

(iii) Part C: Generalization to novel nouns. 18 new Ns, $9\#VX, 9\#CX$. Task was to combine each N with one of the 3 adjs.

(25) Results from Part B (learning A-N phrases):[2] Better performance in N than U conditions (Schane et al., 1974, Figure 1).



a. Not clear what counted as *in*correct, but it couldn't have been restricted to failures to apply the right rule, since chance performance was < 50%.

b. Not shown on graph: Participants learning the U corpus gave more missing responses (38% vs. 23% for UR vs. NR, 38% vs. 15% for US vs. NS).

c. Individuals' curves showed that half of the U group made *increasing* numbers of "natural" responses during the early trials, while the N group got continually better.

[2]Results from Part C are not broken down by N vs. U.

3 Case study: Stress (Carpenter, 2005, 2006)

(26) Attested vs. unattested stress system: "Stress leftmost low vowel, else leftmost" vs. "Stress leftmost high vowel, else leftmost".

a. Stimulus words made by concatenating 3 or 4 isolated CV syllables. Vowels [i u æa] edited to equate duration and intensity. Stressed/unstressed made by manipulating duration, pitch, and intensity.

b. Infrequent stress patterns presented more often than chance to insure enough information to get rule.

c. Pre-test, AXB, to make sure they could hear stressed/unstressed. Criterion is 70% correct. (French speakers got extra training in perception of stress.)

d. Alternate training and testing on complicated schedule.

e. Test on 48 new test words, no feedback.

f. Two participant groups: English speakers in U.S., French speakers in Canada.

g. Results: Proportion correct on novel words (chance = 0.50):

L1	Leftmost low		Leftmost high
English	0.70	$>$	0.62
French	0.59	$>$	0.52

(27) All very well, BUT Carpenter did another experiment, comparing "Stress leftmost heavy syllable, else leftmost" vs. "stress leftmost light syllable, else leftmost".

a. Heavy were CVC, light were CV.

b. Results: Proportion correct on novel words (chance = 0.50):

L1	Leftmost heavy		Leftmost light
English	0.61	=	0.62
French	0.59	=	0.62

4 Other negative reports

(28) Seidl and Buckley (2005, Exp. 1): Familiarized nine-month-old L1 English-learning infants with $C_1V_1C_2V_2(C_3)$ pseudowords, using two patterns of dependency between consonant position and permitted segment:

	Positional restriction		Mean looking time		
Condition	Stops	Fric/Affr	Conforming	|	Nonconforming
"Natural"	C_1	C_2	4.98	$<$	5.75
"Unnatural"	C_2	C_1	5.29	$<$	6.09

a. In familiarization, the stops were [t d b p] and the fricatives/affricates were [s z tʃ dʒ] . In test, some items also contained [g k] and [f v] .

b. No difference between the two conditions.

c. Is there any reason to expect one? The "Natural" condition is supposed to be analogous to intervocalic spirantization. IVS often creates fricatives, but doesn't create affricates. In natural language, intervocalic affricates can become stops or fricatives, rather than the reverse (Kirchner, 1998, 116–118).

(29) Finley and Badecker (pearb): English L1 adults learning backness/rounding harmony.

a. Typological asymmetry: Kaun (1997, §3.1) found the following patterns (plus two more, where trigger and target have to agree in height, as in Yokuts):

	Target		
Trigger	[+high]	[–high]	both
[+high]	Kachin, Hixkaryana, Tsou, Warlpiri, Nyangumata		
[–high]		Mongolian, Tungusic lgs, Murut	
both	Turkic lgs, Nawuri, So. Paiute, Sierra Miwok		Kirgiz

Should we expect generalization from (both, [+high]) to (both, [–high])? What about the reverse?

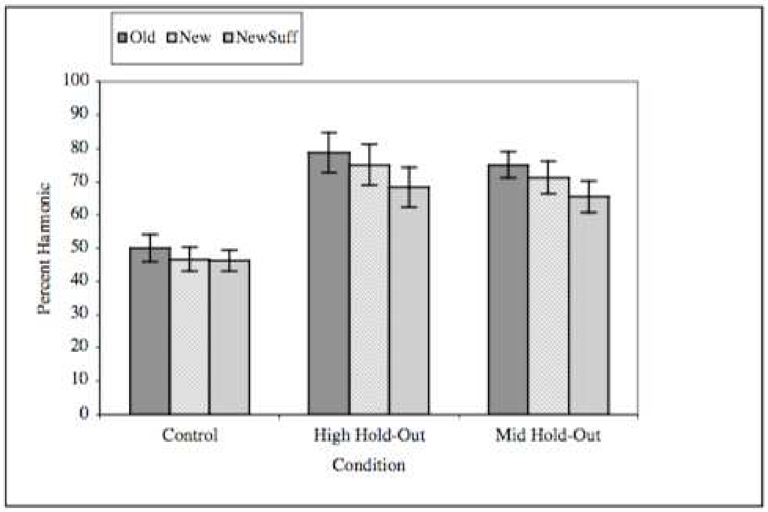
[Rationale: If analytic bias resembles the typology, then Ss trained on (both, [–high]) will tend to infer the Kirgiz pattern, and hence will extend height harmony when introduced to [+high] suffixes. Those trained on (both, [+high]) could infer either the Kirgiz pattern or the Nawuri pattern, and hence might not extend height harmony when introduced to [–high] suffixes.]

b. Train using suffix of one height, test on suffix of a different height. Control condition (no suffix, so learning is impossible) to check untrained bias.

	Training		
Condition	Trigger (stem)	Target (affix)	Test
Mid Hold-Out	both	[+high] bodo-mu	[–high] bodo-ge vs. bodo-go
High Hold-Out	both	[–high] bodo-go	[+high] bodo-gi vs. bodo-gu
Control	Stem only bodo, bido		both random

c. Training was by passive, audio-only exposure to conforming stimuli (24 Stem—Stem+Suffix pairs in the critical conditions, 48 isolated Stems in the Control condition). Stems were $CVCV$. Stems in critical conditions had same vowel twice (one of [i e o u]) so as not to provide direct evidence that harmony ignores height.

d. Test was 36-item 2AFC, with harmonic vs. disharmonic (1/3 old stem + old suffix, 1/3 new stem + old suffix, 1/3 ?? + new suffix).



e. ⇒ No difference between the critical conditions; both patterns learned and generalized equally well.

(30) Omitted for reasons of spacetime: Finley and Badecker (peara); Jusczyk et al. (2002).

5 Discussion

(31) How strong is the evidence for substantive analytic bias? Do phonological learners treat featurally-isomorphic, statistically-equivalent training data differently depending on the real-world interpretation of the features?

(32) Can the positive reports be explained away by learning from L1?

(33) There is a considerable literature on what some have called "hidden rankings", in which speakers distinguish (in perception, production, or intuition) between two different categories of stimuli which are *both* outside their L1 experience (Pertz and Bever, 1975; Davidson et al., 2004; Davidson, 2006; Berent et al., 2007, 2008). In particular, illegal word onsets can be distinguished by their sonority profile, such that some are more illegal than others. Is this evidence of substantive analytic bias?

(34) What made-up experimental result would be conclusive evidence for the existence of substantive analytic bias? (The Data Genie grants you three wishes. . . .)

(35) Suppose substantive analytic bias exists. What can we conclude about its contribution, or lack thereof, to typology?

References

Baayen, R. H., R. Piepenbrock, and L. Gulikers (1995). The CELEX lexical database. CD–ROM, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, Pennsylvania.

Berent, I., T. Lennertz, J. Jun, M. A. Moreno, and P. Smolensky (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences, U.S.A. 105*(14), 5321–5325.

Berent, I., D. Steriade, T. Lennertz, and V. Vaknin (2007). What we know about what we have never heard: evidence from perceptual illusions. *Cognition 104*, 591–630.

Blevins, J. (2006a). Reply to commentaries. *Theoretical Linguistics 32*(2), 245–256.

Blevins, J. (2006b). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics 32*(2), 117–165.

Buckley, E. (2000). On the naturalness of unnatural rules. In *Proceedings from the Second Workshop on American Indigenous Languages*, Volume 9 of *UCSB Working Papers in Linguistics*.

Campbell, L. (2004). *Historical linguistics: an introduction* (2nd ed.). Cambridge, Massachusetts: MIT Press.

Carpenter, A. C. (2005). Acquisition of a natural vs. an unnatural stress system. In A. Brugos, M. R. Clark-Cotton, and S. Han (Eds.), *Papers from the 29th Boston University Conference on Language Development (BUCLD 29)*, Somerville, pp. 134–143. Cascadilla Press.

Carpenter, A. C. (2006). *Acquisition of a natural versus an unnatural stress system*. Ph. D. thesis, University of Massachusetts, Amherst.

Chang, S. S., M. C. Plauché, and J. J. Ohala (2001). Markedness and consonant confusion asymmetries. In E. Hume and K. Johnson (Eds.), *The Role of Speech Perception in Phonology*. San Diego: Academic Press.

Crowley, T. (1997). *An introduction to historical linguistics* (3rd ed.). Oxford, U.K.: Oxford University Press.

Davidson, L. (2006). Phonology, phonetics, or frequency: influences on the production of non-native sequences. *Journal of Phonetics 34*(1), 104–137.

Davidson, L., P. W. Jusczyk, and P. Smolensky (2004). The initial and final states: theoretical implications and experimental explorations of Richness of the Base. In R. Kager, J. Pater, and W. Zonneveld (Eds.), *Constraints in phonological acquisition*, Chapter 10, pp. 321–368. Cambridge, England: Cambridge University Press.

Finley, S. and W. Badecker (to appeara). Right-to-left biases for vowel harmony: evidence from artificial grammar. In *Proceedings of the 38th Meeting of the North-East Linguistics Society*, pp. ??–??

Finley, S. and W. Badecker (to appearb). Towards a substantively biased theory of learning. In *Proceedings of the 33rd Meeting of the Berkeley Linguistics Society*, pp. ??–?? Berkeley Linguistics Society.

Guion, S. G. (1996). *Velar palatalization: coarticulation, perception, and sound change*. Ph. D. thesis, University of Texas, Austin.

Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica 55*, 18–52.

Hale, M. and C. A. Reiss (2000). 'Substance abuse' and 'dysfunctionalism': current trends in phonology. *Linguistic Inquiry 31*(1), 157–169.

Hock, H. H. (1991). *Principles of historical linguistics* (2nd ed.). Berlin: Mouton de Gruyter.

Jusczyk, P. W., P. Smolensky, and T. Allocco (2002). How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition 10*(1), 31–73.

Kaun, A. (1997). *The typology of rounding harmony: an Optimality Theoretic approach*. Ph. D. thesis, University of California, Los Angeles.

Kenstowicz, M. and C. Kisseberth (1979). *Generative phonology: description and theory*. San Diego: Academic Press.

Kiparsky, P. (2006). The amphichronic program vs. Evolutionary Phonology. *Theoretical Linguistics 32*(2), 217–236.

Kirchner, R. (1998). *An effort-based approach to consonant lenition*. Ph. D. thesis, University of California, Los Angeles.

LaRiviere, C., H. Winitz, J. Reeds, and E. Herriman (1974). The conceptual reality of selected distinctive features. *Journal of Speech and Hearing Research 17*(1), 122–133.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: problems and perspectives*, pp. 237–278. Harlow: Longman.

Pertz, D. L. and T. G. Bever (1975). Sensitivity to phonological universals in children and adults. *Language 51*, 149–162.

Schane, S. A., B. Tranel, and H. Lane (1974). On the psychological reality of a natural rule of syllable structure. *Cognition 3*(4), 351–358.

Seidl, A. and E. Buckley (2005). On the learning of arbitrary phonological rules. *Language Learning and Development 1*(3 & 4), 289–316.

Steriade, D. (1997). Phonetics in phonology: the cae of laryngeal neutralization. MS, Department of Linguistics, University of California, Los Angeles.

Steriade, D. (2008). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In S. Inkelas and K. Khanson (Eds.), *The nature of the word: essays in honor of Paul Kiparsky*, Chapter 7, pp. 151–181. Stanford, California: MIT Press.

Trask, R. L. (1996). *Historical linguistics*. London: Arnold.

Wilson, C. (2001). Consonant cluster neutralisation and targeted constraints. *Phonology 18*, 147–197.

Wilson, C. (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science 30*(5), 945–982.

Yu, A. C. L. (2004). Explaining final obstruent voicing in Lezgian: phonetics and history. *Language 80*, 73–97.