NELS 43
CUNY

Pastry phonotactics: Is phonological learning special?

2012 Oct. 19 (F)

Elliott Moreton and Katya Pertsova[1]
University of North Carolina, Chapel Hill (USA)

1 Introduction

(1) Knowledge of language is largely knowledge of featurally-defined patterns, and language learning is, to a great extent, pattern learning. Main research questions of this project (a collaboration with Joe Pater of UMass-Amherst):

a. How does the formal structure of a linguistic pattern affect its learnability in the lab?

b. Does pattern structure affect learning alike or differently across the phonological, morphological, and non-linguistic domains?

c. What implications do formal-structure effects have for the architecture of learning models?

d. Does pattern structure affect learning in the lab the same way it affects typological frequency across natural languages?

(2) Formally identical patterns can occur in different domains both within linguistics and outside it. Here two logical features are instantiated by pairs of phonological, morphological, and visual features:

a. *Phonology*

	Consonant	
Vowel	short	long
short	*lam	$\boxed{\text{lamm}}$
long	$\boxed{\text{laːm}}$	*laːmm

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

b. *Morphology*

	Number	
Case	sing.	pl.
Acc.	mur	$\boxed{\text{mur-s}}$
Nom.	$\boxed{\text{mur-s}}$	mur

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

c. *Non-linguistic game*

	Shapes	
Colors	One	Many
One	Illegal	$\boxed{\text{Legal}}$
Many	$\boxed{\text{Legal}}$	Illegal

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

(3) Outline of this talk:

§2 describes the Shepard hierarchy, a series of patterns on 1 to 3 binary features that has been extensively studied in the non-linguistic pattern-learning literature.

§3 Experiment 1: The Shepard hierarchy is not replicated in a typical "artificial language learning" experiment.

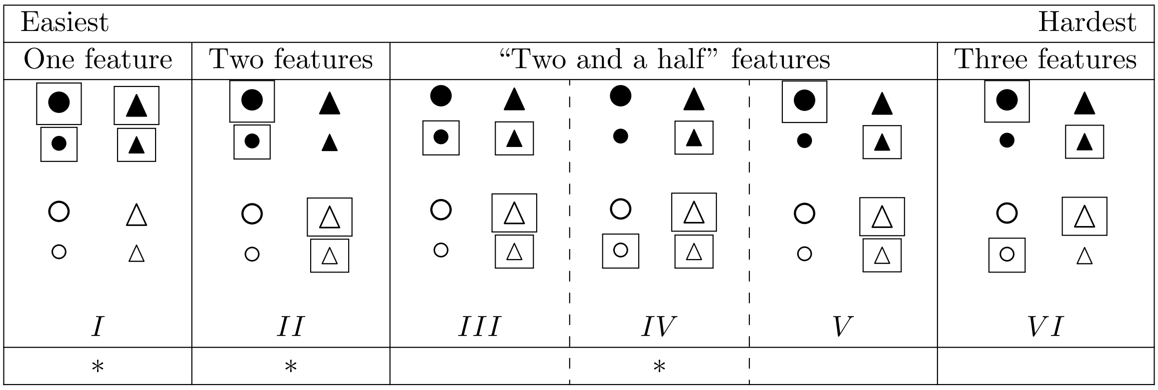
§4 Experiment 2 (in progress, 61%): When the Experiment 1 stimuli are replaced by closely-matched visual analogues, the results are pretty much the same.

§5 Discussion: How do these results relate to the questions we started with?

[1]The work reported here is part of a collaboration with Joe Pater of UMass-Amherst. It has has benefited from discussions with many colleagues, especially Jen Smith (UNC-Chapel Hill), and from audiences at CLS and the Manchester Phonology Meeting earlier this year. We are grateful to Jessica Slavic for assistance with subject-running. All errors are the exclusive responsibility of the authors.

2 The Shepard hierarchy

(4) One structural effect on pattern difficulty that has been extensively studied in the psychology literature is a hierarchy of increasing difficulty for patterns defined on three logical dimensions, instantiated here by the visual dimensions of color, shape, and size (Shepard et al., 1961).



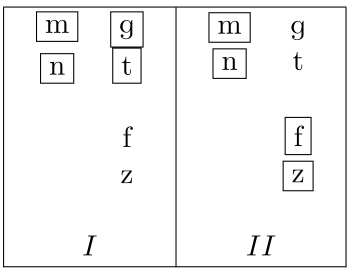
(5) The difficulty order $I > II > \{III, IV, V\} > VI$ has been replicated many times in supervised learning of visual categories (Shepard et al., 1961; Neisser and Weene, 1962; Nosofsky et al., 1994; Feldman, 2000; Love, 2002; Smith et al., 2004), and models of general pattern learning are often evaluated on their ability to reproduce it (Gluck and Bower, 1988a; Anderson, 1991; Kruschke, 1992; Nosofsky et al., 1994; Love et al., 2004; Feldman, 2006).

(6) The easier Shepard types are also more frequent, compared to a chance model, than the harder ones in the Mielke (2008)'s P-Base database of "phonologically active classes". (See Appendix for explanation.)

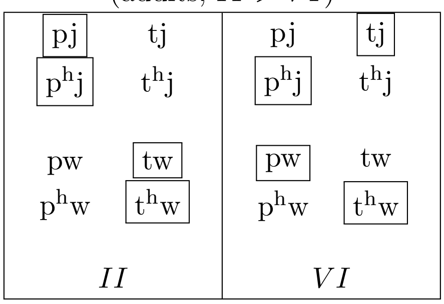
		I	II	III	IV	V	VI
[+syll]	Orig.	840	216	439	197	133	3
(V)	Res.	79	52	322	110	251	8
	Ratio	10.63	4.15	1.36	1.79	0.52	0.38
[–syll]	Orig.	2469	878	3909	2202	2857	79
(C)	Res.	107	100	725	379	604	35
	Ratio	23.07	8.78	5.39	5.81	4.73	2.26

(7) Previous phonological studies have compared Type I with Type II, or Type II with Type VI. They have invariably replicated the $I > II$ and $II > VI$ orders (see Moreton and Pater 2012 for a review). Examples:

Cristiá and Seidl (2008)
(infants, $I > II$)



Kuo (2009)
(adults, $II > VI$)



(8) Although Experiments 1 and 2 address all six types, this talk will focus on Types I, II, and IV (starred in 4 above):

Type I: Only one feature matters.

Type II: Two correlated features.

Type IV: “Family resemblance” (prototype) structure.

---

3 Experiment 1: Pattern structure and difficulty in phonological learning

(9) Experiment 1 compared all 6 types using a typical “artificial-language” methodology.

a. *Stimuli*: MBROLA-synthesized $C_1V_1C_2V_2$ words with inventory /t k d g/ /i u æ ɔ/, used previously by Moreton (2008); Lin (2009); Kapatsinski (2011); Moreton (2012). There were 256 possible words.

Feature	Stimulus segment			
	σ_1		σ_2	
	C_1	V_1	C_2	V_2
voiced	±		±	
Coronal	±		±	
high		±		±
back		±		±

Consonants				Vowels			
k	t	g	d	æ	ɔ	i	u
−	−	+	+				
−	+	−	+				
				−	−	+	+
				−	+	−	+

b. *Participants*: 141 (out of a planned 144) paid volunteers from the UNC-Chapel Hill community, self-screened for normal hearing and native English. Each participant was randomly assigned to one of Types I–VI (24 people per Type).

c. *Patterns*: For each participant, 3 of the 8 stimulus features were randomly chosen, then randomly mapped onto the 3 logical features defining the Shepard type to define a “language”. Examples:

TYPE I: C1 is voiced **d**igu, **g**ada, **d**ika, **g**ugu, . . .
TYPE II: C1 is voiced iff V2 is back. **d**ig**u**, **t**æg**i**, **k**ag**æ** **g**ad**a**, . . .
TYPE IV: At least two of: C1 is voiced, V2 is high, V2 is back **k**ak**u**, **d**ig**u**, **g**uk**i**, **d**æk**a**, . . .

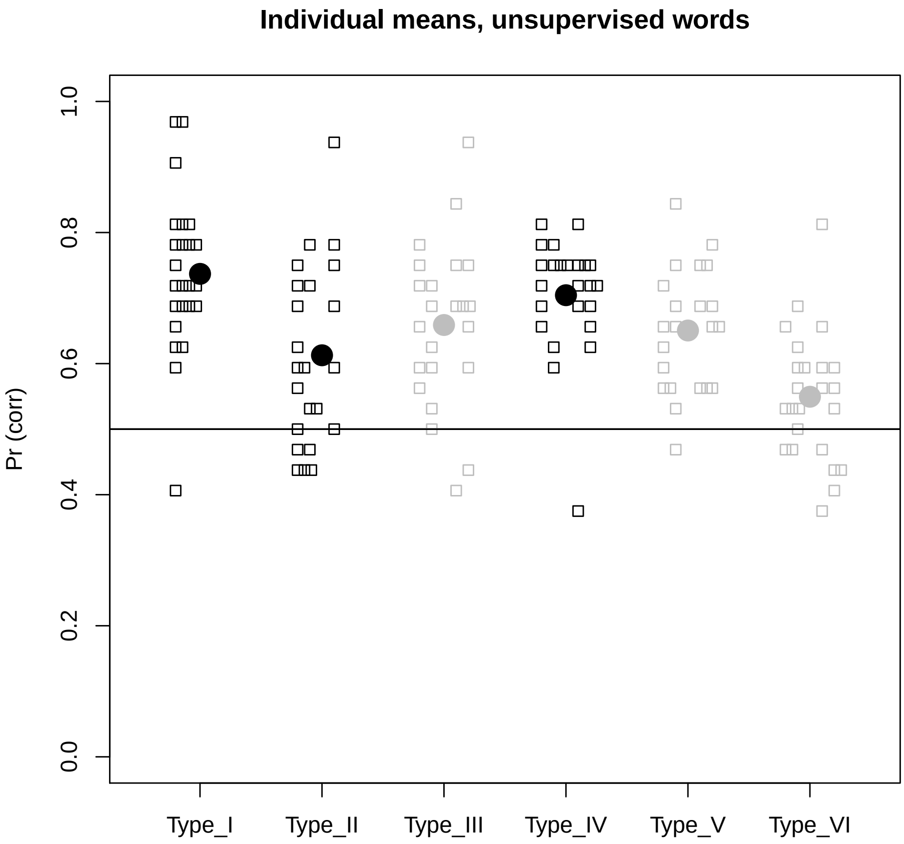
Since patterns were randomly generated without regard to typological frequency or phonetic motivation, they were very likely to be “crazy rules” (Bach and Harms, 1972; Anderson, 1981).

d. *Instructions*: Participants were told they would learn to pronounce words in an artificial language, and then be tested on ability to recognize words in that language.

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

f. *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”.

(10) Results for Types I, II, and IV. Each plotting symbol represents one participant.



(11) Analysis by mixed-effects logistic regression with *Participant* as a random effect and the following fixed effects:

a. *Type*, I, II, IV, with I as the reference category

b. Nuisance variables *Reduplicated* and *FirstInPair*, explained in Moreton (2012).

(12) The fixed-effects part of the fitted model is shown below. Type I is the reference category.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.98948	0.12616	7.843	4.40e-15	***
Type_II	-0.62149	0.16479	-3.771	0.000162	***
Type_IV	-0.19034	0.16520	-1.152	0.249254	
FirstInPair	0.20454	0.09254	2.210	0.027084	*
Reduplicated	-0.65904	0.14156	-4.655	3.23e-06	***

(13) The fitted model was used to test the $I/II/IV$ contrasts, with significance levels adjusted for multiple simultaneous comparisons (Hothorn et al., 2008). Results, compared to classic Shepard results:

Classic Shepard

	II	IV
I	$>$	$>$
II	—	$>$

Experiment 1

	II	IV
I	$> ***$	n.s.
II	—	$< *$

(14) **Interim summary**: Pattern structure affected learning in this experiment, but not in the same way that it does in the classic Shepard experiments. In particular, the $II > IV$ advantage was reversed in Experiment 1.

(Compare similar results of Pertsova 2012 on morphological vs. non-linguistic pattern learning.)

4 Experiment 2: Non-phonological analogues

(15) Q. Do the results of Experiment 1 mean that phonological learning is inherently different from visual category learning?

A. Not necessarily. There are findings in the literature that other factors can cause the classic $II > IV$ advantage to disappear or reverse even in visual pattern learning (Nosofsky and Palmeri, 1996; Love, 2002; Smith et al., 2004; Kurtz et al., 2012).

(16) Exp. 1 was designed to be like other "artificial-language" experiment, which in turn are designed to be like natural-language learning — and which therefore differ from the classic Shepard experiments in several ways.

Classic Shepard	Phonological learning (incl. Exp. 1)		Exp. 2
Visual domain	Phonological domain		Visual domain
Easily verbalizable features ("red triangle")	Features hard for naive participants to verbalize ("voiceless velar")		Easily verbalizable features ("pink icing")
Overt instructions to learn a pattern	No mention of pattern in instructions		No mention of pattern in instructions
Positive and negative examples used in training; corrective feedback on every trial ("supervised" learning)	Only positive examples; no feedback ("unsupervised" learning)		Only positive examples; no feedback ("unsupervised" learning)
3 features	8 features, of which 3 are critical and 5 are distractors		8 features, of which 3 are critical and 5 are distractors
No within-stimulus structure	Stimuli have internal prosodic and feature-tier structure		Stimuli have analogues of prosodic and feature-tier structure

(17) ⇒ An alternative explanation for the II/IV reversal is that participants could reason explicitly about the features of the visual stimuli used in the Shepard experiments, but not about phonological features. Previous research shows that explicit learning can favor Type II over Type IV (Love, 2002; Smith et al., 2004; Kurtz et al., 2012).

(18) Experiment 2 asks whether the II/IV reversal observed in Experiment 1 is reduced or eliminated when the phonological stimuli are replaced by closely analogous visual stimuli, which can be reasoned about explicitly.

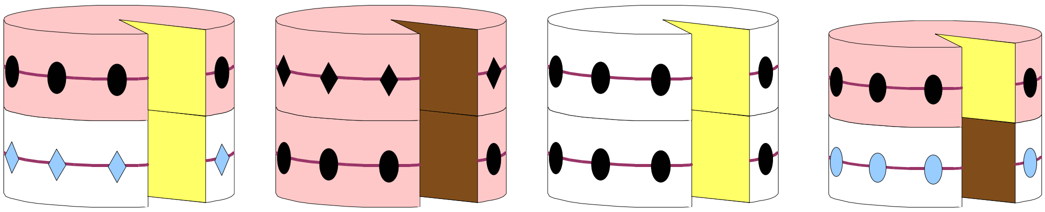
(For previous work on visual analogues of artificial-language experiments, see Finley and Badecker 2010.)

(19) Stimuli were 8-feature fancy cakes, organized into layers ($\simeq$ syllables) and body vs. decoration ($\simeq$ vowels and consonants):

Feature	Stimulus segment			
	σ_1		σ_2	
	C_1	V_1	C_2	V_2
voiced	±		±	
Coronal	±		±	
high		±		±
back		±		±

Feature	Nonlinguistic analogues			
	Layer 1 (Bottom)		Layer 2 (Top)	
	Decoration 1	Body 1	Decoration 2	Body 2
Diamond candy	±		±	
Blue candy	±		±	
White icing		±		±
Brown batter		±		±

(20) Each of the 256 possible stimulus words thus has an corresponding cake:

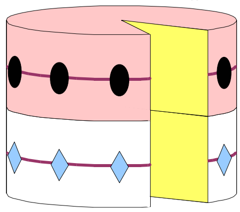


[dikæ] [kɔgɔ] [kiki] [tugæ]

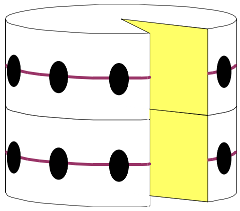
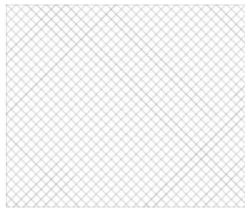
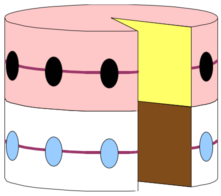
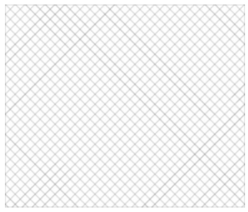
(21) Differences between Experiment 2 and Experiment 1:

a. *Instructions*: Participants were told that they would be learning to recognize "a particular style of fancy cake". They would first study cakes made in this style, then they would be "tested on how well you can recognize them."

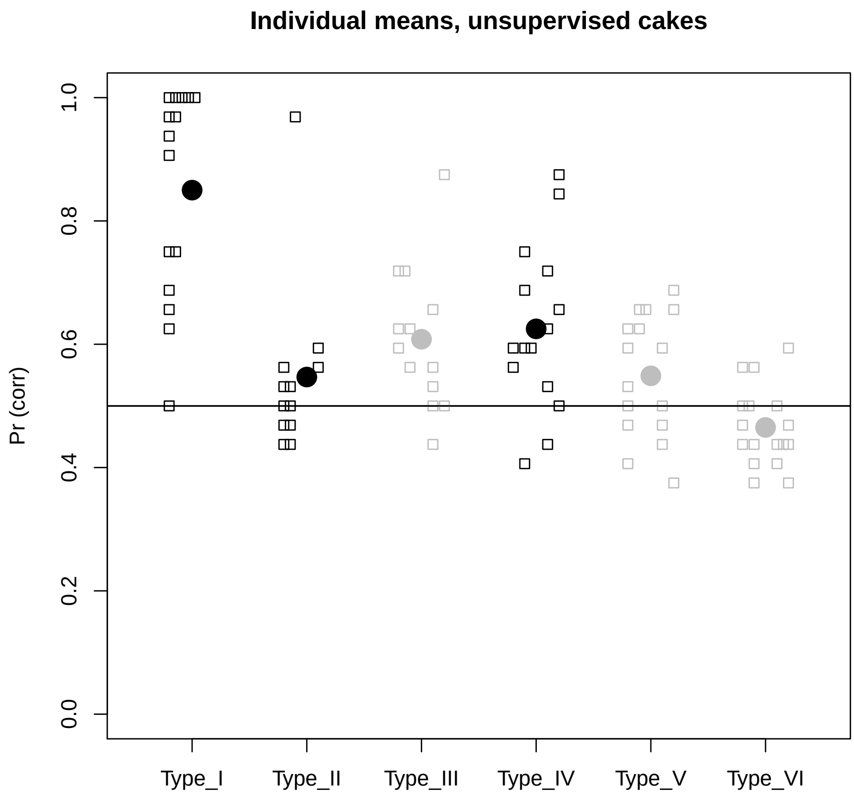
b. *Familiarization*: Participants viewed 32 pattern-conforming cakes in random order 4 times. They could look at each cake as long as they liked.

Visual mask	Blank screen	Positive example
		
250 ms	250 ms	$>$ 2000 ms

c. *Test*: 32 two-alternative forced-choice trials, each with one new pattern-conforming cake and one non-conforming cake:

Fixation point + 1000 ms	Blank screen 250 ms	2AFC option  2000 ms	Visual mask  250 ms
Blank screen 250 ms	2AFC option  2000 ms	Visual mask  250 ms	Blank screen > 250 ms

(22) Interim results ($N = 88$ out of 144)



(23) Interim analysis (61% of planned data), same model as for Experiment 1. Fixed-effects part of model:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     2.0724     0.2890   7.172 7.40e-13 ***
Type_II        -1.9533     0.3976  -4.913 8.96e-07 ***
Type_IV        -1.6286     0.3778  -4.311 1.63e-05 ***
FirstInPair     0.2507     0.1273   1.970   0.0489 *
Reduplication  -0.2740     0.1934  -1.417   0.1566
```

(24) Comparison to typical Shepard results (significance levels adjusted for multiple comparisons):

Classic Shepard

	II	IV
I	$>$	$>$
II	—	$>$

Experiment 1

	II	IV
I	$> ***$	n.s.
II	—	$< *$

Experiment 2 (61%)

	II	IV
I	$> ***$	$> ***$
II	—	n.s.

In Experiment 1, Type IV was no harder than Type I, and was easier than Type II. In Experiment 2, Type IV is (so far) harder than Type I and not easier than Type II.

(25) Interim summary: Switching to more-verbalizable visual stimuli has not (so far) restored Type II to being easier than Type IV.

⇒ The isomorphic word and cake experiments seem to be engaging similar kinds of learning in both domains.

5 Discussion

(26) The research questions from p. 1 again:

a. How does the formal structure of a linguistic pattern affect its learnability in the lab?

b. Does pattern structure affect learning alike or differently across the phonological, morphological, and non-linguistic domains?

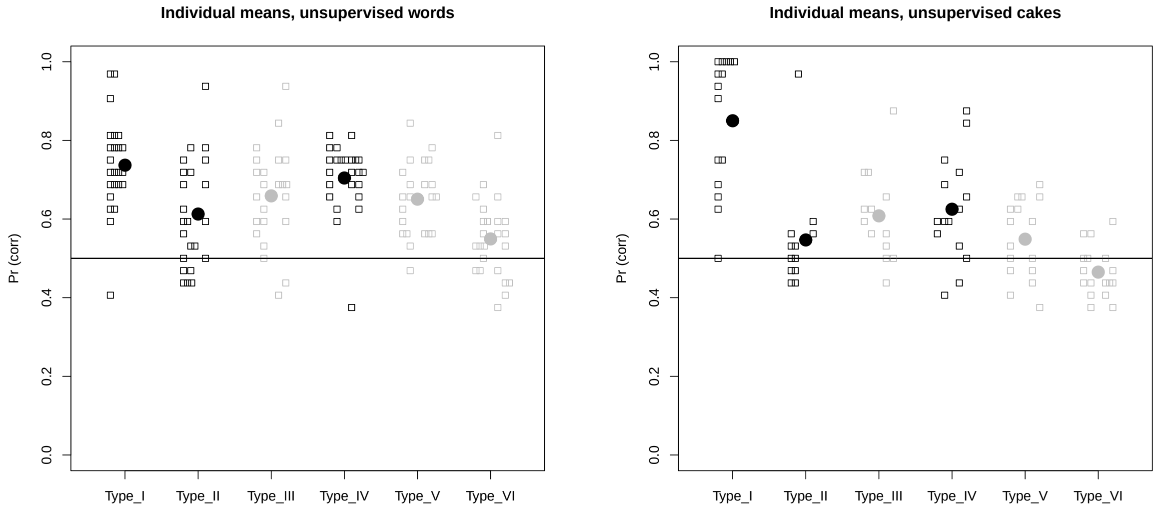
c. What implications do formal-structure effects have for the architecture of learning models?

d. Does pattern structure affect learning in the lab the same way it affects typological frequency across natural languages?

5.1 Is phonological learning special?

(27) Is phonological learning special? Are the effects of pattern structure on learning difficulty different in phonology and other domains?

(28) Experiment 1 showed that the order of pattern difficulty in a typical "artificial-language" experiment differs from the order $I > II > III, IV, V > VI$ found in the classic experiments with non-linguistic categories. Experiment 2 (in progress) substitutes closely analogous visual stimuli for the phonological stimuli of Experiment 1, but this manipulation does not restore the classic Shepard order.



⇒ When isomorphic phonological and visual stimuli are used, with similar instructions and experimental paradigm, *pattern structure affects learning similarly in both domains.*

5.2 Implications for modelling

(29) If phonological learning isn't special, then phonologists can exploit the vast body of empirical and theoretical work in psychology on pattern learning, and vice versa.

(30) The results of Experiments 1 and 2 pose a challenge for models of pattern learning that are designed to reproduce the Shepard hierarchy (e..g., the "rational model" (Anderson, 1991), ALCOVE (Kruschke, 1992), RULEX (Nosofsky et al., 1994), the Boolean complexity model (Feldman, 2000), SUSTAIN (Love et al., 2004).)

These models are also architecturally very different from most phonological learning models.

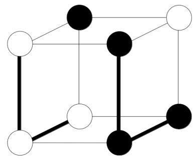
(31) More promising: The Incremental Maximum Entropy with a Conjunctive Constraint Schema (IMECCS, Pater and Moreton 2012; Pater 2012) learner is an adaptation to phonology of the Configural Cue Model of Gluck and Bower (1988b,a) — which was rejected as a model of non-linguistic category learning precisely because it overestimated the difficulty of Type II. Main properties:

a. Unbiased conjunctive constraints: One for every conjunction of +, −, or "don't care" feature values over the whole stimulus set ($[-F_2], [+F_2 - F_3], [+F_1 + F_2 - F_3], \ldots]$).

b. Error-driven learning: On each training trial, adjust the influence of a constraint up or down in proportion to its effect on the output error. Algorithms of this sort exist for Stochastic OT (Boersma, 1997; Boersma and Hayes, 2001; Magri, 2008), Harmonic Grammar (Pater, 2008; Boersma and Pater, 2008), and Maximum Entropy grammar (Jäger, 2007).

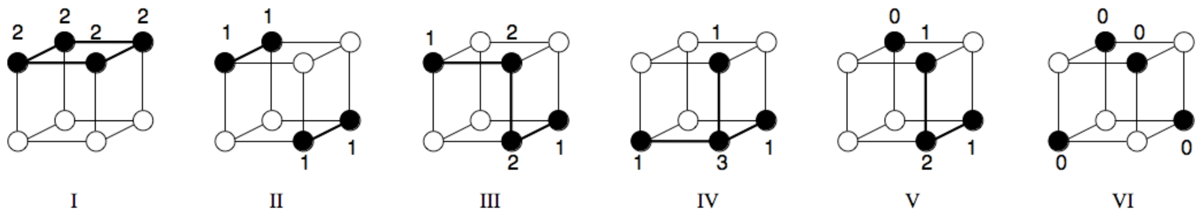
(32) IMECCS predicts that after training, classification behavior will be largely determined by what Gluck and Bower (1988a, 188–189) called "partially valid cues", i.e., constraints that apply to a subset of the data and that correctly classify that subset.

(33) A constraint can be valid by favoring positive stimuli, or by disfavoring negative ones. For example, Shepard Type V has four valid two-feature constraints, corresponding to the bolded cube edges in this figure:



These two-feature constraints will turn out to be so important that we will give them a name, *valid edges.*

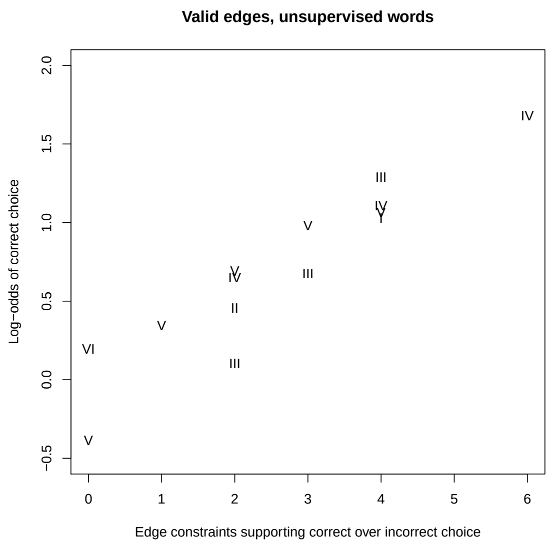
(34) Within each concept, the individual stimuli are supported by different numbers of valid edges. This illustration shows only how many support each positive stimulus; the negative stimuli are symmetrical.



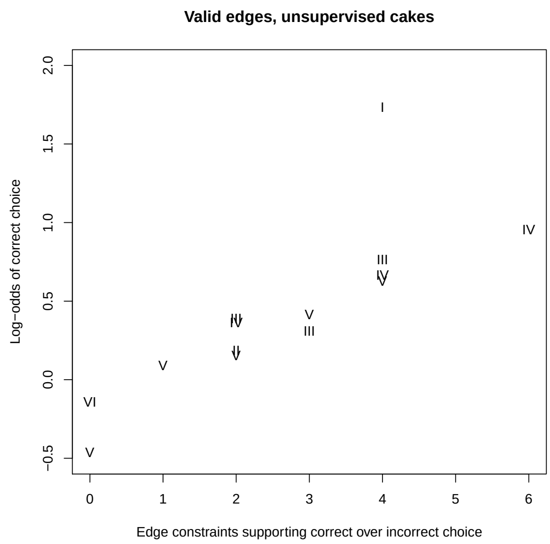
(35) On each trial of the experiment, the participant choses between one positive and one negative stimulus. The total number of valid edges supporting the positive over the negative stimulus can range from 0 to 6.

(36) Results of Experiments 1 and 2, replotted to show log-odds of a correct response as a function of how many valid edges favored the positive over the negative stimulus:

Experiment 1



Experiment 2



⇒ Difficulty of a test pair is determined by the number of valid edges, as predicted by IMECCS.

(Except for Type I in Experiment 2, which is unexpectedly easy — perhaps because participants can reason explicitly about single features of cakes.)

5.3 Pattern structure and typology

(37) Typological frequency may be causally related to learning difficulty: The harder something is to learn, the more likely it is to be lost over time, or changed to something different and easier (Bach and Harms, 1972).

⇒ Whatever biases phonological learning is subject to, we might see their effects in disparities between the actual typological frequencies of different pattern types and the frequencies expected by chance.

(38) The easier Shepard types are also more frequent, compared to a chance model, than the harder ones in the Mielke (2008)'s P-Base database of "phonologically active classes". (See Appendix for explanation.)

		I	II	III	IV	V	VI
[+syll]	Orig.	840	216	439	197	133	3
(V)	Res.	79	52	322	110	251	8
	Ratio	10.63	4.15	1.36	1.79	0.52	0.38
[–syll]	Orig.	2469	878	3909	2202	2857	79
(C)	Res.	107	100	725	379	604	35
	Ratio	23.07	8.78	5.39	5.81	4.73	2.26

(39) This typological sample agrees with the classic Shepard results, and *dis*agrees with the phonological learning results of Experiment 1! How could this be?

a. P-Base describes natural classes, and Experiment 1 was about phonotactic dependencies, so they are not "about" the same thing.

b. Lab experiments differ from real L1 or L2 acquisition circumstances in many ways (shorter, smaller, less semantic,, supervised learning, etc.). Maybe our lab experiments aren't yet good enough simulations of natural acquisition.

5.4 Summary

(40) **Main point**: Phonological learning in the lab is affected by pattern structure in much the same way as visual learning of analogous stimuli. (I.e., these experiments do not show that phonological learning is special.)

⇒ Psychological models and experimental results related to non-linguistic category learning may be applicable as well to phonological learning, and vice versa.

(41) Some questions for the future:

a. But is phonological learning special? There are many other structural effects; do *they* differ across domains?

   Or can all apparently special properties of phonological learning be traced back to the unusual structure of the stimulus space (e.g., prosodic and autosegmental relations)?

b. Why does pattern structure seem to affect phonotactic learning differently from natural-class typological frequency?

c. What about learning in linguistic domains where the features are easier to verbalize, such as morphology? Should Type II be easier than Type IV? If so, how must models of morphological learning differ from models of phonological learning?

References

Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review 98*, 409–429.

Anderson, S. R. (1981). Why phonology isn't "natural". *Linguistic Inquiry 12*, 493–539.

Bach, E. and R. T. Harms (1972). How do languages get crazy rules? In R. P. Stockwell and R. K. S. Macaulay (Eds.), *Linguistic change and generative theory*, Chapter 1, pp. 1–21. Bloomington: Indiana University Press.

Boersma, P. (1997). Functional Optimality Theory. *Proceedings of the Institute of Phonetic Sciences of University of Amsterdam 21*, 37–42.

Boersma, P. and B. Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry 32*, 45–86.

Boersma, P. and J. Pater (2008, May). Convergence properties of a gradual learning algoritm for Harmonic Grammar. MS.

Chomsky, N. and M. A. Halle (1968). *The sound pattern of English.* Cambridge, Massachusetts: MIT Press.

Cristiá, A. and A. Seidl (2008). Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development 4*(3), 203–227.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature 407*, 630–633.

Feldman, J. (2006). An algebra of human concept learning. *Journal of mathematical psychology 50*, 339–368.

Finley, S. and W. Badecker (2010). Linguistic and non-linguistic influences on learning biases for vowel harmony. In S. Ohlsson and R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Sciety*, Austin, Texas, pp. 706–711. Cognitive Science Society.

Gluck, M. A. and G. H. Bower (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language 27*, 166–195.

Gluck, M. A. and G. H. Bower (1988b). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General 117*, 227–247.

Hothorn, T., F. Bretz, and P. Westfall (2008). Simultaneous inference in general parametric models. *Biometrical Journal 50*(3), 346–363.

Jäger, G. (2007). Maximum Entropy models and Stochastic Optimality Theory. In J. Grimshaw, J. Maling, C. Manning, J. Simpson, and A. Zaenen (Eds.), *Architectures, rules, and preferences: a festschrift for Joan Bresnan*, pp. 467–479. Stanford, California: CSLI Publications.

Kapatsinski, V. (2011). Modularity in the channel: the link between separability of features and learnability of dependencies between them. Proceedings of the XVIIth International Congress of Phonetic Sciences.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review 99*, 22–44.

Kuo, L. (2009). The role of natural class features in the acquisition of phonotactic regularities. *Journal of psycholinguistic research 38*(2), 129–150.

Kurtz, K. J., K. R. Levering, R. D. Stanton, J. Romero, and S. N. Morris (2012). Human learning of elemental category structures: revising the classic result of Shepard, Hovland, and Jenkins (1961). *Journal of Experimental Psychology: Learning, Memory, and Cognition xx*(xx), xxxx–xxxx.

Lin, Y. (2009). Tests of analytic bias in native Mandarin speakers and native Southern Min speakers. In Y. Xiao (Ed.), *21st North American Conference on Chinese Linguistics*, Smithfield, Rhode Island, pp. 81–92. Bryant University.

Löfstedt, I. (1992). Swedish segment length and Structure Preservation. *Studia Linguistica 46*(2), 93–127.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin and Review 9*(4), 829–835.

Love, B. C., D. L. Medin, and T. M. Gureckis (2004). SUSTAIN: a network model of category learning. *Psychological Review 111*(2), 309–332.

Luquiens, F. B. (1909). *An introduction to Old French phonology and morphology.* New Haven, Connecticut: Yale University Press.

Magri, G. (2008). Linear methods in Optimality Theory: a convergent incremental algorithm that performs both promotion and demotion. MS, Department of Linguistics and Philosophy, Massachusetts Institute of Technology.

Mielke, J. (2004). *The emergence of distinctive features.* Ph. D. thesis, Ohio State University.

Mielke, J. (2008). *The emergence of distinctive features.* Oxford, England: Oxford University Press.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology 25*(1), 83–127.

Moreton, E. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal*

of Memory and Language 67(1), 165–183.
Moreton, E. and J. Pater (2012, April). Structure and substance in artificial phonology learning. Manuscript, University of North Carolina at Chapel Hill and University of Massachusetts at Amherst.
Neisser, U. and P. Weene (1962). Hierarchies in concept attainment. *Journal of Experimental Psychology 64*(6), 640–645.
Nosofsky, R. M., M. A. Gluck, T. J. Palmeri, S. C. McKinley, and P. Gauthier (1994). Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition 22*(3), 352–369.
Nosofsky, R. M. and T. J. Palmeri (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin and Review 3*(2), 222–226.
Nosofsky, R. M., T. J. Palmeri, and S. C. McKinley (1994). Rule-plus-exception model of classification learning. *Psychological Review 101*(1), 53–79.
Pater, J. (2008). Gradual learning and convergence. *Linguistic Inquiry 39*(2), 334–345.
Pater, J. (2012, October 6). Incremental Maximum Entropy phonotactics and the Shepard complexity hierarchy. Handout from a presentation at the North-East Conference on Phonology (NECPhon), University of Maryland.
Pater, J. and E. Moreton (2012). Structurally biased phonology: complexity in learning and typology. *Journal of the English and Foreign Languages University, Hyderabad*.
Pertsova, K. (2012). Logical complexity in morphological learning. To appear in *Proceedings of the Berkeley Linguistics Society*.
Ross, S. M. (2006). Qwirkle: mix, match, score, and win! Game instructions, MindWare Company.
Shepard, R. N., C. L. Hovland, and H. M. Jenkins (1961). Learning and memorization of classifications. *Psychological Monographs 75*(13, Whole No. 517).
Smith, J. D., J. P. Minda, and D. A. Washburn (2004). Category learning in rhesus monkeys: a study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General 133*(3), 398–404.

A Appendix: The Shepard types in phonological typology

(42) Does formal structure affect typological frequency the same way it affects phonological learning in the lab? I.e., are easier pattern types also more frequent?

(43) Empirical data from `P-base1.93` (Mielke, 2008), which contains 627 `Language` entries and 9041 classes of `Segments`. Entries look like this:

```
Language,Maori
Reference,Harlow, Ray (1996) M?ori. Muenchen: Lincom Europa.
Family,AUSTRONESIAN
Location,New Zealand
Langcode,MAOR
Inventory,Core
p,t,k,f,h,i,u,m,n,?,e,o,?,a,w,
Inventory,Marginal

Trigger,/t/ ? palatalized / __X
Segments,i,u,
Maybe,
```

a. Standardized minor formatting irregularities in files, but did not change language data.

b. Excluded languages if they contained anything unexpected, such as apparent double occurrences of the same segment in inventory statements. This left 620 languages (99%) of the original database.

c. Excluded classes if their descriptions contained anything unexpected, such as double occurrences of same segment. This left 8971 classes (again 99% of the original database).

(44) Separate analysis for consonants and vowels: The analysis was restricted to classes whose members all had the same value of [+syllabic]. This eliminated 1262 classes (14.1%), leaving 2034 [+syllabic] and 5682 [–syllabic] classes.

(45) Standard of comparison: Over- or underrepresented relative to *what*? Suppose we find that 40% of patterns can be expressed as Type II, and only 20% as Type III. Does that mean that languages somehow "favor" Type II over Type III?

(46) The comparison we really want to make is between the typology we actually have, and the one we would expect if learners were unbiased between Shepard classes. That would involve knowing precisely what other typologically-effective factors would be left if we took out the learning biases (e.g., the structure and magnitude of phonetic precursors, etc.) — which we don't know at all.

Next best thing: We created simulated classes following the procedure used by Mielke (2004, 194): For each of the 2034 [+syllabic] and 5682 [–syllabic] classes in P-Base, a new class of the same size was created by randomly sampling the [+syllabic] or [–syllabic] sub-inventory with uniform probability.

(47) Each "phonologically active class" in both the original and the resampled P-Base was processed as follows to obtain all of the logical structures it was consistent with.

a. We used P-Base's *SPE* feature system (Chomsky and Halle, 1968) because it uses the standard [±high] and [±low] for vowel height.

b. An *expression* is an assignment of phonetic feature to logical features that allows the positive class members to be distinguished from the negative ones using only the logical features. (I.e., no smallest cell contains both a positive and a negative class member.) Example:

Class 734, Expression 277019, Unami Delaware

	[–voice]		∼[–voice]	
	[–distr]	∼ [–distr]	[–distr]	∼ [–distr]
[–cont]	t	p, tʃ, k	$\boxed{\text{n}}$	$\boxed{\text{m}}$
∼[–cont]	$\boxed{\text{s}}$	$\boxed{\text{ʃ, x, h}}$	ˑ	w,j

c. Three-feature expressions were found for 1780 of the 2034 [+syllabic] classes (87.5%) and 3501 of the 5682 [–syllabic] ones (61.5%) in the original P-base. In the resampled one, the corresponding numbers were 1251/2034 (61.6%) and 1136/5682 (20.0%).

(48) Each expression which had four positive and four negative cells was then processed to see which Shepard type it fell into. The results are shown in this table.

		I	II	III	IV	V	VI
[+syll]	Orig.	334	2	0	1	0	0
(V)	Res.	8	4	18	5	33	3
[–syll]	Orig.	219	4	7	4	3	0
(C)	Res.	0	2	2	0	2	0

a. Distinct classes belonging to at least one Shepard type: [+syll] original 337, resampled 71; [+syll] original 236, resampled 6.

b. Type I is massively overrepresented in the original P-Base compared to the resampled P-Base.

(49) To get more resolution of the higher types, we investigated "defective" expressions, i.e., those where the inventory of the language left some cells empty:

Class 6, Expression 2850. Sakhalin Ainu

	[–voice]		∼[–voice]	
	[+cons]	∼ [+cons]	[+cons]	∼ [+cons]
[–cont]	p,t,tʃ,k	ʔ	$\boxed{\text{m,n}}$	
∼[–cont]	$\boxed{\text{s}}$	$\boxed{\text{h}}$	r	$\boxed{\text{w,j}}$

In some cases, the empty cells allowed unambiguous coercion to a Shepard pattern. (E.g., if only one cell of the 8 is empty, there is only one way to fill it in to create a Shepard pattern.)

(50) Each expression was processed to see what Shepard types it was consistent with. Expressions which were consistent with multiple Shepard types were discarded. The remaining expressions were used to count the number of classes that had an expression that was consistent with each Shepard type. Results:

		I	II	III	IV	V	VI
[+syll]	Orig.	840	216	439	197	133	3
(V)	Res.	79	52	322	110	251	8
	Ratio	10.63	4.15	1.36	1.79	0.52	0.38
[–syll]	Orig.	2469	878	3909	2202	2857	79
(C)	Res.	107	100	725	379	604	35
	Ratio	23.07	8.78	5.39	5.81	4.73	2.26

The original/resampled odds ratios decline as Shepard type increases, for both vowel and consonant classes. Here are the pairwise significant differences for the probability that a class is from original P-Base, given that it can be represented as a particular Shepard type (2-sample test with continuity correction, and significance levels corrected for multiple simultaneous comparisons):

[+syll] (V)

	II	III	IV	V	VI
I	>***	>***	>***	>***	?
II	–	>***	>***	>***	?
III	–	–	<marg.	>***	?
IV	–	–	–	>***	?
V	–	–	–	–	?

[–syll] (C)

	II	III	IV	V	VI
I	>***	>***	>***	>***	>***
II	–	>***	>***	>***	>***
III	–	–	<***	>***	>***
IV	–	–	–	>***	>***
V	–	–	–	–	>marg.

(51) *Caution*: The natural-language classes are classes of segments, i.e., the relevant features all occur in the same segment. In the phonological-learning experiment, the relevant features could occur anywhere in the *CVCV* stimulus word. Thus, the experiment was not a perfect analogue of the typological survey.