Pastry phonotactics: Is phonological learning special?

Elliott Moreton & Katya Pertsova

University of North Carolina at Chapel Hill

1. Introduction: Language learning as pattern learning

Knowledge of language is largely knowledge of featurally-defined patterns, and language learning is, to a great extent, pattern learning. A simple example is shown in (1), where the same logical pattern is instantiated by different pairs of phonological, morphological, and visual features. The recognition of this commonalty invites many questions. How does the formal structure of a linguistic pattern affect its learnability in the lab? Does pattern structure affect learning alike or differently in different inductive domains, such as morphological, phonological, and non-linguistic patterns? What implications do formal-structure effects have for the architecture of learning models? Does pattern structure affect learning in the lab the same way it affects typological frequency across natural languages?

(1) Instantiation of the same exclusive-or pattern in three domains.

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*

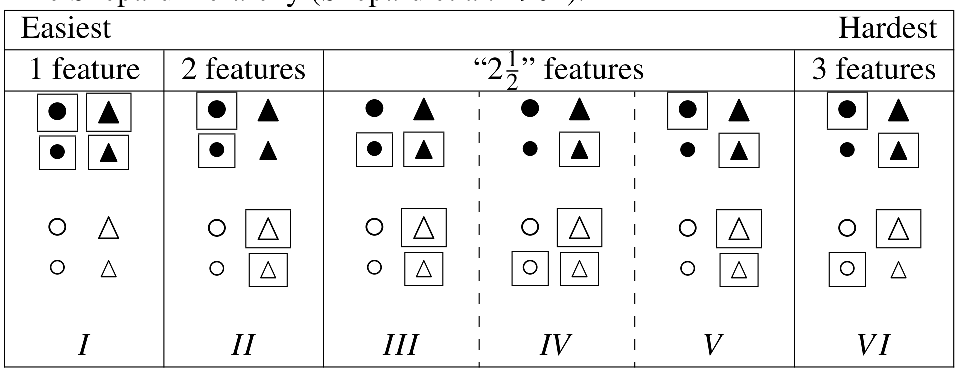
	Shapes	
Colors	One	Many
One	Illegal	$\boxed{\text{Legal}}$
Many	$\boxed{\text{Legal}}$	Illegal

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

*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 the Chicago Linguistic Society and the Manchester Phonology Meeting in 2012. We are grateful to Jessica Slavic for assistance with subject-running. All errors are the exclusive responsibility of the authors. This work was funded in part by a grant from UNC-Chapel Hill to the second author.

We are interested in the hypothesis that linguistic patterns which differ in formal structure also differ systematically in how hard they are to learn. This hypothesis has been extensively studied by psychologists interested in the learning of patterns defined on non-linguistic features. One finding of these studies has been a hierarchy of increasing difficulty for a particular set of patterns defined on three logical dimensions. The hierarchy is shown in (2), instantiated by the visual dimensions of color, shape, and size.

(2) The Shepard hierarchy (Shepard et al. 1961).



The difficulty order $I > II > \{III, IV, V\} > VI$ has been partially or wholly replicated many times in supervised learning of visual categories (Shepard et al. 1961, Nosofsky et al. 1994, Feldman 2000, Love 2002, Smith et al. 2004), and competing models of general pattern learning are often evaluated on their ability to reproduce it (Anderson 1991, Kruschke 1992, Nosofsky et al. 1994, Love et al. 2004, Feldman 2006). The most successful models, by this measure, have been *rule-based* models in which the learner tests discrete rule-like hypotheses in increasing order of the number of crucial features used in each hypothesis (Nosofsky et al. 1994, Goodman et al. 2008). *Cue-based* models, which learn by gradually updating weights on pre-specified property detectors, have been unsuccessful in matching the Shepard order. In particular, they incorrectly predict Type IV to be easier than Type II, because they are sensitive to whether a category is linearly separable in feature space (Medin and Schwanenflugel 1981, Gluck and Bower 1988).

Observed difficulty order can therefore be informative about which kind of learning processes the learner is using. In this paper, we are particularly interested in the hypothesis that phonological learners are using cue-based processes, because cue-based models that are very similar to psychological models have been independently proposed in linguistics for phonological learning (e.g., Boersma 1998, Goldwater and Johnson 2003, Jäger 2007, Hayes and Wilson 2008, Pater and Moreton 2012). For this reason, we focus on three empirical questions:

1. Is the classical Shepard order ($I > II > \{III, IV, V\} > VI$) also found in a typical short-term phonological-learning ("artificial-language") experiment? (Preview: It is not; in particular, Type II turns out to be harder than Types III and IV.)

2. Is the Shepard order restored when the phonological stimuli are replaced with visual analogues? (Preview: It is not; Type II is still harder than higher types)

3. Is the Shepard order found in natural-language phonological patterns? (Preview: It is.)

These findings support the hypothesis that phonological learning is cue-based, and that it shares this property with some kinds of visual learning. They invite further inquiry into the relation between phonological learning in the lab and typological frequency in natural languages, and into the conditions favoring rule-like vs. cue-like learning (Kurtz et al. 2013).

2. Experiment 1: Pattern structure and difficulty in phonological learning

Previous studies of phonological learning 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 2012a,b for a review). Experiment 1 compared all 6 types using a typical "artificial-language" methodology. Stimuli were MBROLA-synthesized $C_1V_1C_2V_2$ words with inventory /t k d g/ /i u æ ɔ/, used previously by Moreton (2012). There were 256 possible words, as shown in (3).

(3) Stimulus design for Experiment 1.

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

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

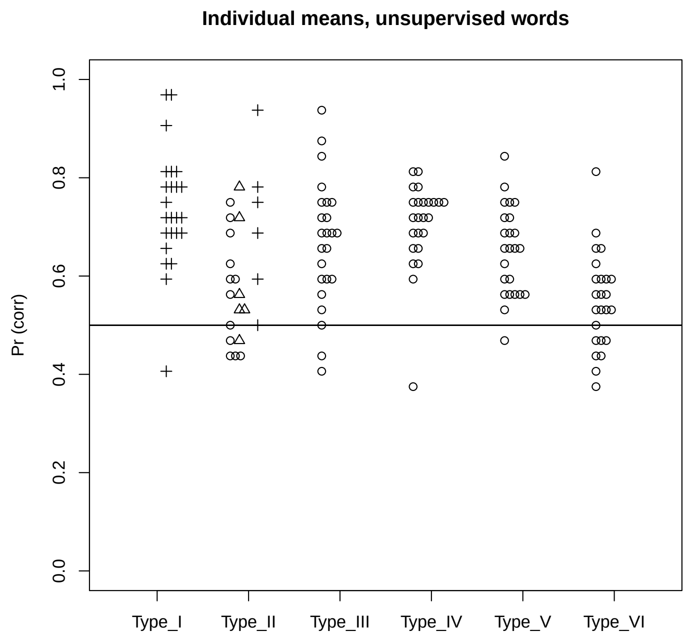
Participants were 144 paid volunteers from the authors' university community, self-screened for normal hearing and native English. (An additional 8 participants were dropped from the study because of native language, equipment failure, or failure to meet a criterion of at least 10 correct responses.) Each participant was randomly assigned to one of Types I–VI (24 participants per Type). 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 are shown in (4). Since the patterns were randomly generated, with deliberate disregard of typological frequency and phonetic motivation, they were almost sure to be "crazy rules" (Bach and Harms 1972, Anderson 1981). This was intentional, since the focus of the study is *purely structural* effects on phonological learning.

(4) Instantiation of Shepard patterns in phonological stimuli: 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 kak**u**, **d**ig**u**, **g**uk**i**, **d**æk**a**, ...

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. They were familiarized with the pattern by listening to and repeating aloud 32 randomly-chosen pattern-conforming stimuli 4 times over. Then, in the test phase, 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".

(5) Individual participant results from Experiment 1. Each plotting symbol represents one participant ($N = 144$). The vertical axis is the proportion of "correct" (i.e., pattern-conforming) choices in the test phase. Plotting symbols are explained in text.



Results, shown in (5), clearly do not conform to the classic Shepard difficulty hierarchy. Performance on Type II patterns is, on average, *worse* than on Types III, IV, and V. The overall numerical order of the means for the types is $I > IV > III > V > II > VI$. The statistical analysis, shown in (6), used Type II as the reference category. The other Type conditions were dummy coded (e.g., for a participant in the Type III condition, the factor `III` was 1, and the variables `I, IV, V`, and `VI` were 0).

Type II patterns involved only two features, and it could happen that both of those features belonged to the same feature genus (e.g., they were both voicing features), so that the pattern was harmony or disharmony. Such patterns are often easier to learn than other Type II patterns (see review in Section 3.1 of Moreton and Pater 2012a), so a factor `AllSameGenus` was set to 1 for the 6 harmony/disharmony patterns (plotted with a "△" in (5)), 0 for all others.

Another factor, `AllSameSeg`, was set to 1 for patterns in which all of the critical features occurred in the same segment, i.e., all of the Type I patterns and 6 of the Type II patterns (plotted with a "+" in (5)). Finally, two nuisance factors, `Redup` and `CorrFirst`, were included to model out variance caused by participants' aversion to reduplicated stimuli and their preference for the first of the two-alternative forced-choice stimuli (Moreton 2008, 2012). A mixed-effects logistic-regression model was fitted with the `lmer` function of the *lme4* package in R 2.7.1.

The analysis confirmed that Type II performance was significantly below performance on Types III, IV, and V, and in fact did not significantly exceed performance on Type VI. Thus, pattern structure affected learning in this experiment, but not in the same way that it does in the classic Shepard experiments.

(6) Fixed-effects part of the mixed-effects logistic-regression model of participant data from Experiment 1 (4608 responses from 144 participants; log-likelihood = -2879). Type II is the reference category.

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.12800  0.14337   0.893 0.371990
I             0.12644  0.23434   0.540 0.589497
III           0.47203  0.17323   2.725 0.006432 **
IV            0.63402  0.17423   3.639 0.000274 ***
V             0.38985  0.17266   2.258 0.023951 *
VI           -0.06131  0.17119  -0.358 0.720248
AllSameSeg    0.69359  0.25150   2.758 0.005819 **
AllSameGenus  0.11085  0.24330   0.456 0.648652
Redup        -0.77233  0.10142  -7.615 2.63e-14 ***
CorrFirst     0.27397  0.06348   4.316 1.59e-05 ***
```

3. Experiment 2: Visual analogues

Do the results of Experiment 1 mean that phonological learning is inherently different from visual category learning? 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. 2013). Experiment 1 was designed to be like other “artificial-language” experiments, which in turn are designed to be like natural-language learning — and which therefore differ from the classic Shepard experiments in several ways, as shown in (7).

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 conditions that favor explicit learning can favor Type II over Type IV (Love 2002, Smith et al. 2004, Kurtz et al. 2013). 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, Lai 2012.)

Our stimuli were 8-feature fancy cakes, organized into layers ($\simeq$ syllables) and body vs. decoration ($\simeq$ vowels and consonants), as shown in (8). Each of the 256 possible stimulus words thus has an corresponding cake, as shown in (9).

(7) Differences between Exp. 1, Exp. 2, and the classic Shepard-like experiments

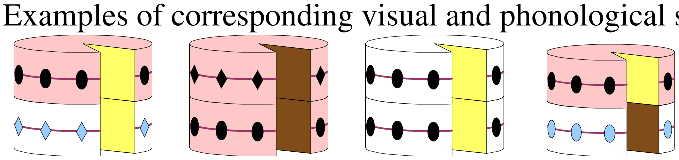
Classic Shepard	Phonological learning (incl. Exp. 1)	Exp. 2
Visual domain	Phonological domain	Visual domain
Easily verbalizable features (“red triangle”)	Features hard for naïve participants to verbalize (“voiceless velar”)	Easily verbalizable features (“pink icing”)
Overt instructions to learn a pattern	No mention of pattern in instructions	No mention of pattern in instructions
Supervised learning	Unsupervised learning	Unsupervised learning
3 features, all critical	8 features, 3 critical and 5 distractors	8 features, 3 are critical and 5 distractors
No within-stimulus structure	Stimuli have internal prosodic and feature-tier structure	Stimuli have analogues of prosodic and feature-tier structure

(8) Design of visual stimuli for Experiment 2, showing isomorphism with phonological stimuli of Experiment 1.

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

	Nonlinguistic analogues			
	Bottom layer		Top layer	
Feature	Candy	Body	Candy	Body
Diamond candy	±		±	
Blue candy	±		±	
White icing		±		±
Brown batter		±		±

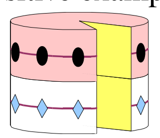
(9) Examples of corresponding visual and phonological stimuli in Exps. 1 and 2.



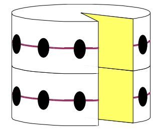
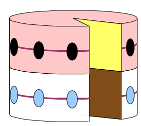
[dikæ] [kɔgɔ] [kiki] [tugæ]

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.” In the familiarization phase, participants viewed 32 pattern-conforming cakes in random order 4 times (see 10). They could look at each cake as long as they liked before going on to the next one. The test phase consisted of 32 two-alternative forced-choice trials, each with one new pattern-conforming cake and one non-conforming cake (see 11). There were 144 participants from the same population as in Exp. 1 (9 more were dropped due to equipment failure, four for native language or failure to meet criterion).

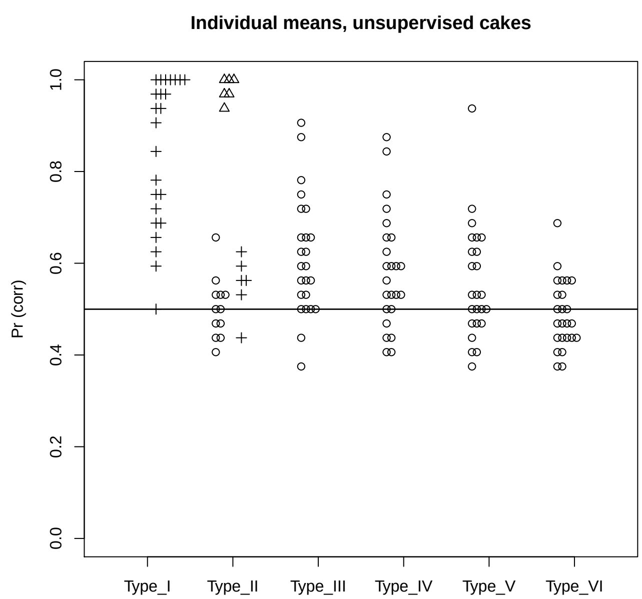
(10) Familiarization phase in Experiment 2

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

(11) Test phase in Experiment 2

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

(12) Individual participant results from Exp. 2. Each point represents one participant ($N = 144$). The vertical axis is the proportion of pattern-conforming choices in the test phase. Plotting symbols are explained above, at (5).



(13) Fixed-effects part of the mixed-effects logistic-regression model of participant data from Experiment 2 (4608 responses from 144 participants; log-likelihood = −2798). Type II is the reference category.

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.17460    0.17077  -1.022   0.3066
I             1.63189    0.27667   5.898 3.67e-09 ***
III           0.48629    0.20646   2.355   0.0185 *
IV            0.35651    0.20607   1.730   0.0836 .
V             0.24296    0.20586   1.180   0.2379
VI           -0.05399    0.20544  -0.263   0.7927
AllSameGenus  3.94821    0.58222   6.781 1.19e-11 ***
AllSameSeg    0.21899    0.29093   0.753   0.4516
CorrFirst     0.35884    0.06437   5.574 2.49e-08 ***
Redup        -0.23879    0.09933  -2.404   0.0162 *
```

The question was whether replacing the phonological stimuli with visual analogues would restore the classic Shepard order. Results, plotted in (12), show that it did not. Except for the 6 participants in the harmony/disharmony subcase (who achieved near-perfrct performance), performance in Type II was still no better than in Types III, IV, and V. The statistical analysis, shown in (13), bears this conclusion out: Type II performance was numerically *below* that in Types III, IV, and V, with the comparison reaching significance for Type III.

4. The Shepard hierarchy in phonological classes

In this section of the paper, we ask how structure affects typological frequency, and in particular, whether pattern types which are easier in the lab are also more frequent in nature. We addressed this question using the largest machine-readable database of phonological patterns known to us, `P-base1.93` (Mielke 2008), which contains 627 `Language` entries and 9041 classes of `Segments`. Each entry describes a "phonologically active class", i.e., one which triggers a change, undergoes a change, or figures in a static phonotactic pattern. A typical entry is shown in (14).

We standardized minor formatting irregularities in files, but did not change any language data. Languages were excluded 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. Classes were excluded if their descriptions contained anything unexpected, such as apparent double occurrences of same segment. This left 8971 classes (still 99% of the original database).

We used P-Base's *SPE* feature system (Chomsky and Halle 1968) because it uses the standard [±high] and [±low] for vowel height. However, rather than allow the *SPE* convenience feature [± syllabic] to figure in classes, we carried out separate analyses for consonant and vowel classes: 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.

(14) `P-base1.93` entry describing triggers of a palatalization pattern in Māori.

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

Trigger,/t/ → palatalized / __X
Segments,i,u,                                 ← Class involved in pattern
Maybe,
```

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? 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. Instead, our chance model 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 from the relevant language's [+syllabic] or [–syllabic] sub-inventory with uniform probability.

In order to assign classes to logical types, 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. An *expression* was defined as an assignment of phonological features to logical features that allows the positive class to be distinguished from the rest of the inventory. An example is shown in (15).

(15) Example: Class 734, Expression 277019, Unami Delaware. This way of expressing the class creates four filled positive cells and four filled negative cells, instantiating Shepard Type II. "∼" means "not"; e.g., "∼[–cont]" includes both [+cont] segments and those unspecified for [cont]. These segments are the entire recorded consonant inventory of the language.

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

Most classes yielded multiple expressions. A word of caution is in order here: We do not know which expression, if any, best describes a speaker's mental representation of the class. For example, the class in (15) could also be described as "[+nas] or ([+cont] and [–voice])". *Some* of these expressions could be categorized into Shepard types, and we analyzed those expressions of those types.

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%). 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 (16). Only a minority of the classes could be assigned to any Shepard type at all: 337 [+syllabic] classes and 236 [–syllabic] classes in the original P-Base, and 71 [+syllabic] and 6 [–syllabic] in the resampled one.

(16) Frequency of Shepard classes in original and resampled `P-base1.93` (Mielke 2008).

		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

It is clear that Type I is greatly overrepresented in the original P-Base compared to the resampled chance model. However, Types II–VI are so rare in the original P-Base that there is no way to make comparisons among them. To get more resolution of the higher types, we investigated "defective" expressions, i.e., those where the inventory of the language left some (positive or negative) cells empty, as shown in (17).

(17) Example: Class 6, Expression 2850, Sakhalin Ainu. The language's inventory leaves one (negative) cell empty. The rest are consistent with Shepard Type V.

	[–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.) Each expression was processed to see what Shepard types it was consistent with. If an expression was consistent with multiple Shepard types, it was discarded. For each class and each Shepard type, we then asked whether there was at least one undiscarded expression for that class which was consistent with that Shepard type. Each time the answer to that question was "yes", we counted one more instance of that Shepard type. Thus, a single class could count towards more than one type. A wide variety of feature-based classes was found; i.e., the results were not dominated by a handful of very frequent classes. Summary results are shown in (18).

(18) For each Shepard type, this table shows the number of P-Base classes for which there is at least one expression that is consistent with only that type.

		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

If the original and resampled P-Bases were sampling from the same distribution of types, then when the two same-sized samples are pooled, a class which has a Type II expression (for instance) should be just as likely to come from the original as the resampled P-Base. Clearly, the original P-Base samples from a distribution that produces more Shepard classes, and the original/resampled odds ratios decline as Shepard type increases; i.e., lower-numbered types are *more* overrepresented than higher ones in the original P-Base relative to the resampled one. Degrees of overrepresentation are compared in (19). For example, in the left-hand (vowel) panel, the cell in the row labelled "I" and the column labelled "II" contains the symbol ">" to signify that the ratio of original to resampled P-Base vowel classes of Type I, 840 to 79, is significantly greater than the ratio of original to resampled P-Base vowel classes of Type II, 216 to 52. The significance criterion was $p < 0.001$ by 2-sample exact binomial test, after significance levels were Bonferroni-deflated for multiple simultaneous comparisons.

(19) Pairwise significant differences for the conditional probability that a class is from the original P-Base, based on counts in (18). "–" means no comparison, "?" means no significant difference. See text for methods.

[+syll] (V)

	II	III	IV	V	VI
I	>	>	>	>	?
II	–	>	>	>	?
III	–	–	?	>	?
IV	–	–	–	>	?
V	–	–	–	–	?

[–syll] (C)

	II	III	IV	V	VI
I	>	>	>	>	>
II	–	>	>	>	>
III	–	–	<	>	>
IV	–	–	–	>	>
V	–	–	–	–	?

The typological results are in accordance with the classical Shepard hierarchy: $I > II > III, IV, V$, with the additional refinement that $IV, III > V$ (for vowel classes) or even $IV > III > V$ (for consonant classes). Consequently, they *mis*match the results of Experiment 1 in an important way: If the typological rarity of a pattern type were directly predictable from experimental difficulty, then Type II ought to be less frequent than Types III, IV, and V, rather than more frequent.

5. Discussion

Exp. 1 showed that, in a typical short-term phonological-learning experiment, the logical structure of a pattern affects participants' uptake of it. Performance on Type II patterns was significantly below that on Types III, IV, and V, unlike a large body of results in visual pattern learning. Exp. 1 is consistent with cue-based learning (Gluck and Bower 1988), and so is compatible with some independently-motivated phonological-learning models (Boersma 1998, Goldwater and Johnson 2003, Hayes and Wilson 2008, Pater and Moreton 2012). In Exp. 2, similar results were obtained with visual analogues; hence, we cannot say that Exp. 1 revealed a peculiarity unique to phonological learning. On the other hand, the P-Base survey found that Type II patterns exceeded chance frequency by significantly more than did patterns of Types III, IV, and V. What are we to make of this? There are several possibilities; here are three that strike us as interesting and testable.

Possibility #1: P-Base and Exp. 1 are incomparable. The P-Base classes are segment classes, i.e., the relevant features all occur in the same segment, while Exp. 1, the relevant features could occur anywhere in the *CVCV* stimulus word. However, part of Exp. 1 did compare Types I and II when all relevant features co-occured in the same segment, and found no significant difference between Types I and II (this was the subcase where `AllSameSeg` was 1; see discussion of (6) above). This possibility can be tested with single-segment Shepard experiments that are more closely analogous to the P-Base survey, or by doing typological surveys that code inter-segmental dependencies.

Possibility #2: Phonetics is structurally biased. P-Base tabulates classes that developed in real languages exposed to the effects of phonetic channel bias, and which may therefore be phonologizations of phonetic precursors (for a review, see Hansson 2008). Perhaps the precursors are biased, i.e., phonetic variables may tend to covary in ways that resemble continuous analogues of Types I and II rather than Types III and IV. This possibility awaits research in phonetic typology.

Possibility #3: Natural language is influenced by some other kind of learning. Short-term adult phonological-learning experiments may simulate early stages of *second*-language learning better than first-languaeg acquisition. Slower processes that escape study in the lab may yet steer acquisition and language change towards Type II and away from the three-feature types. This possibility points towards more-naturalistic experimental studies, observational study of structural effects in acquisition and historical change, and agent-based modelling of inductive- and channel-bias effects on typology.

Comparative study of pattern learning and typology across linguistic and non-linguistic domains may lead to a "Grand Unified Theory", to a recognition that culture of all sorts is the product of special-purpose mechanisms, or to some unexpected alternative.

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. (1998). *Functional Phonology: formalizing the interactions between articulatory and perceptual drives*. Ph. D. thesis, University of Amsterdam.

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

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 (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language 27*, 166–195.

Goldwater, S. J. and M. Johnson (2003). Learning OT constraint rankings using a maximum entropy model. In J. Spenader, A. Erkisson, and O. Dahl (Eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, pp. 111–120.

Goodman, N., J. B. Tenenbaum, J. Feldman, and T. L. Griffiths (2008). A rational analysis of rule-based concept learning. *Cognitive Science 32*(1), 108–154.

Hansson, G. Ó. (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass 2*, 859–893.

Hayes, B. and C. Wilson (2008). A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry 39*(3), 379–440.

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.

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

Kurtz, K. J., K. R. Levering, R. D. Stanton, J. Romero, and S. N. Morris (2013). Human learning of elemental category structures: revising the classic result of Shepard, Hovland, and Jenkins (1961). *Journal of Experimental Psychology: Learning, Memory, and Cognition 39*(2), 552–572.

Lai, Y. R. (2012). *Domain specificity in learning phonology*. Ph. D. thesis, University of Delaware.

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.

Medin, D. L. and P. J. Schwanenflugel (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory 7*(5), 355–368.

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 (2012a). Structure and substance in artificial-phonology learning: Part I, structure. *Language and Linguistics Compass 6*(11), 686–701.

Moreton, E. and J. Pater (2012b). Structure and substance in artificial-phonology learning: Part II, substance. *Language and Linguistics Compass 6*(11), 702–718.

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. and E. Moreton (2012). Structurally biased phonology: complexity in learning and typology. *Journal of the English and Foreign Languages University, Hyderabad 3*(2), 1–44.

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.

Department of Linguistics
University of North Carolina
Chapel Hill, NC 27599–3155

Contact us at moreton@email.unc.edu or pertsova@email.unc.edu