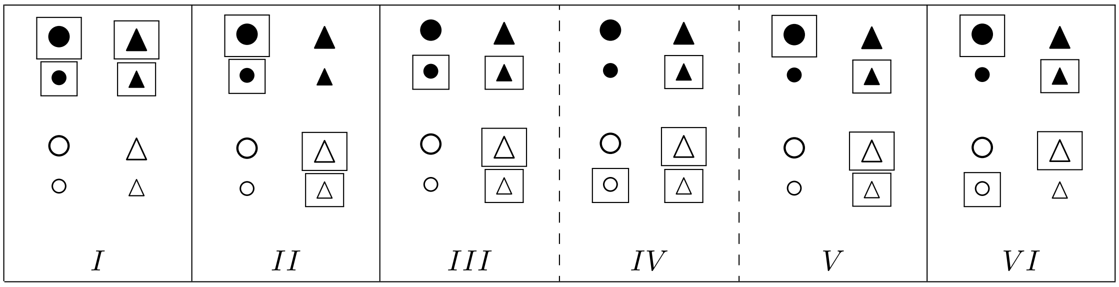
Elliott Moreton & Katya Pertsova
(University of North Carolina at Chapel Hill) **Friday 10:45**

Pastry phonotactics: Is phonological learning special?

Gist: How does the formal structure of a linguistic pattern affect its learnability in the lab and its typological frequency in natural language? Does structure affect learning alike or differently for phonological and non-linguistic patterns? (1) A well-studied difficulty hierarchy for non-linguistic patterns does not predict the typological frequencies of analogous phonological patterns. (2) Pattern difficulty in a phonological learning experiment *does* match the relevant typological facts, while mismatching the non-linguistic learnability hierarchy. (3) A non-linguistic learning experiment (in progress) tests whether this difference persists when the non-linguistic stimuli are given analogues of prosodic and feature-tier structure.

Structural complexity: A hierarchy of complexity and learning difficulty has been established for *non*-linguistic patterns defined on one to three features (Figure 1, after Shepard et al. 1961): Learnability decreases in the order $I > II > \{III \approx IV \approx V\} > VI$. (E.g., the Type I pattern "black figures" is easier than the Type II pattern "black circles plus white triangles".)

Figure 1.



Typological frequency: If phonological learning is affected by complexity in the same way, the harder patterns might be less frequent in natural language. Are they? We analyzed the "phonologically active classes" in P-Base (Mielke, 2008), tabulating the patterns that could be unambiguously assigned to a Shepard type. (E.g., "front rounded vowels plus back unrounded vowels" would be Type II.) Chance expectations were established by randomly resampling P-Base to get an equal-sized database of size- and inventory-matched classes (Mielke, 2004). Frequency relative to chance was quantified as (observed/(observed+expected)).

The relative typological frequencies resulting from the analysis of Mielke's database (Figure 2b) mismatch the usual non-linguistic difficulty order (Figure 2a). Specifically, Types III, IV, and V are of similar difficulty in non-linguistic learning, but differ in relative typological frequency ($IV > III > V$). Type V is also only weakly distinguished from Type VI in typological frequency.[1]

Figure 2.

a. Non-ling. difficulty (from Fig. 1)

	II	III	IV	V	VI
I	>	>	>	>	>
II	–	>	>	>	>
III	–	–	n.s.	n.s.	>
IV	–	–	–	n.s.	>
V	–	–	–	–	>

b. Typological frequency relative to chance

	II	III	IV	V	VI
I	$>^{***}$	$>^{***}$	$>^{***}$	$>^{***}$	$>^{***}$
II	–	$>^{***}$	$>^{***}$	$>^{***}$	$>^{***}$
III	–	–	$<^{***}$	$>^{***}$	$>^{***}$
IV	–	–	–	$>^{***}$	$>^{***}$
V	–	–	–	–	>marg.

Experiment 1, Artificial phonological patterns: This disparity suggests that phonological learning should resemble phonological typology rather than non-linguistic learning. Does it? The hypothesis was tested using a typical "artificial-language" paradigm. Each participant ($N = 140$ to date) was assigned a type (I–VI). A phonotactic pattern of that type was chosen, with the 3 relevant features randomly selected from among the height, backness, voicing, and place (Cor vs. Dor) features of a set of 256 $CVCV$ tokens. For example: The random procedure might choose [place of C_1], [backness of V_1], and [voicing of C_2] to play the roles played by

[1] 2-sample exact binomial test, significance levels corrected for multiple simultaneous comparisons.

color, shape, and size in Figure 1, with a further random assignment dictating that [C_1 = Cor] corresponds to black, [V_1 = −back] to circular, and [C_2 = +voiced] to large. A phonological Type I pattern would then oppose stimuli like [dagu], [tiki], [dugæ], etc. to stimuli like [kudi], [kægu], [giku], etc. Phonological patterns of Types II through VI were constructed analogously. The participant practiced pronouncing 32 random pattern-conforming stimuli 4 times. They then heard 32 new conforming/nonconforming pairs and tried to pick the conforming one.

Results are shown in Figure 3a.[2] Phonological pattern learning, like phonological typology but unlike previous non-linguistic pattern-learning results, distinguishes among Types III, IV, and V. Type IV is so easy as to be indistinguishable from Type I, whereas Type V is so hard as to be indistinguishable from Type VI.

Figure 3.

a. Phonological pattern (N = 140 participants)

	II	III	IV	V	VI
I	>***	>*	n.s	>***	>***
II	–	n.s.	n.s.	n.s.	> **
III	–	–	n.s.	>marg.	>***
IV	–	–	–	>***	>***
V	–	–	–	–	n.s.

b. Cake style (N = 40 participants)

	II	III	IV	V	VI
I	>***	> *	> ***	>*	>***
II	–	n.s.	n.s.	n.s.	n.s.
III	–	–	n.s.	n.s.	>**
IV	–	–	–	n.s.	> (marg.)
V	–	–	–	–	> *

Experiment 2, Isomorphic non-linguistic pattern learning: These results seem to favor the hypothesis that phonological learning and typology show different complexity effects from non-linguistic learning. However, previous experiments on non-linguistic learning differ from Experiment 1 in major ways: (A) They use supervised learning (explicit training with right/wrong feedback) and (B) their stimuli do not have the characteristic internal structure of phonological stimuli — prosodic and feature-tier organization. Perhaps it is these factors that make the difference, rather than phonological vs. non-linguistic domain.

To control for these factors, an experiment is in progress which is (A) unsupervised, like Experiment 1, and (B) uses non-linguistic stimuli that are close analogues of the phonological stimuli of Experiment 1: fancy cakes. For each stimulus word, there is a corresponding cake. Syllables correspond to layers, and phonological features correspond to properties of the batter, icing, and decorations (Figure 4 shows corresponding words and cakes). Instead of a fictitious language, participants learn to recognize a fictitious style of cake. Each random "language" in Experiment 1 is matched by an isomorphic cake style in Experiment 2.

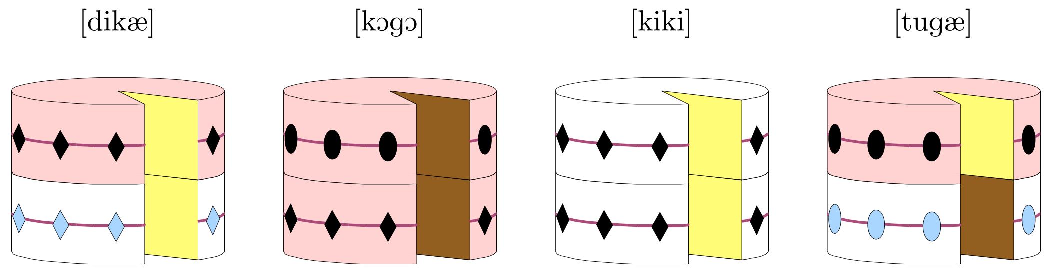


Figure 4.

Early results (N = 40 participants out of a planned 144) are shown in Figure 3b. So far, they resemble the classic non-linguistic results (2a), rather than the phonological results (3a), in that Type IV is significantly harder than Type I and not distinct from Types III and V, while Type V is significantly easier than Type VI. These interim findings support the hypothesis that **there are qualitative differences between pattern learning in phonology and in other domains**.

[2]Analyzed using mixed-effects logistic regression with Type as a between-participant fixed effect, and Participant as a random effect. Pairwise significant differences by Tukey's Honest Significant Difference procedure with 95% family-wise confidence level (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)