DGfS, AG7
Uni Leipzig

The comparative study of inductive bias

2015 March 5 (Θ)

E. Moreton, UNC-CH

1 Introduction

(1) *The research program*: Comparative study of inductive bias in pattern-learning across domains (phonological, morphological, and non-linguistic patterns).

Collaboration with Joe Pater (UMass-Amherst), Katya Pertsova (UNC-Chapel Hill). Thanks to RAs: Rachel Broad, Metta Crouse, Caleb Hicks. (N.B. I'm speaking only for myself today.)

(2) *Inductive bias*: Prejudice in the learner in favor of some hypotheses and against others. Every learner that can generalize has them, as they are needed to choose between hypotheses that aren't distinguished by the data (Pinker, 1979; Mitchell, 1990; Gallistel et al., 1991, e.g.,).

(3) *Structural inductive bias* discriminates between patterns on the basis of the formal relationships between the features, rather than the real-world content of those features. Some examples of proposed structural biases in linguistic theory:

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

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

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

(4) *Comparative study across domains*: Linguistic patterns may be structurally isomorphic to non-linguistic patterns ("concepts"):

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 category*

	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).

[0]The work reported here is part of a continuing collaboration with Joe Pater and Katya Pertsova. I am indebted to several other colleagues for ideas, discussion, and critique, including Ewan Dunbar, Matt Goldrick, Jen Smith, Paul Smolensky, two anonymous *Cognitive Science* reviewers, as well as to audiences at Sound Change 2014 at Berkeley, and at the 2014 LAGB meeting in Oxford. Any remaining errors are mine. Some of the work was funded by an internal grant from the University of North Carolina at Chapel Hill. Address for correspondence about this talk: `moreton@unc.edu`.

The logical *structure* is the same in all three cases (IFF/XOR between two binary-valued features), though the *substance* differs greatly.

(5) *Significance for linguistics*: Why do linguists care about structural inductive bias, and why is it worth comparing across domains?

a. *Typology* emerges from the interaction of inductive bias with other factors (notably phonetic factors that systematically distort the phonological form of the learner's input, Hyman 1976; Ohala 1992, 1993). Together, inductive and other biases skew language change, and through it the long-term steady-state frequencies of patterns (Bell, 1970, 1971; Greenberg, 1978).

b. Inductive biases can be informative about the *algorithmic architecture of the learner*, and hence also about the form of the grammar. (E.g., does it look like a set of weights, or a list of rules?) The study of structural bias in non-linguistic pattern learning has spawned a rich empirical and theoretical literature, which has developed largely in isolation from analogous work in linguistics.

(6) Outline of talk:

a. Architecture and inductive bias: "cue-based" and "rule-based" learners differ.

b. In non-linguistic domains, these architectures have been proposed to align with implicit and explicit learning, respectively.

c. Exp. 1: Are implicit and explicit learning available in phonology too? (Answer: *Yes.*)

d. Exp. 2: Do they differ in structural inductive bias, in the same way they do in non-linguistic learning? (Answer: Partly *yes*, partly *can't tell.*)

e. Discussion

---

2 "Cue-based" and "rule-based" architectures

(7) Focus: Two different model architectures for learning, representing, and using patterns:

a. *Cue-based* models: A large population of predicates participates simultaneously in determining the output. Learning means gradually adjusting their relative importance, and this adjustment can affect any number of predicates at once. Example: The Gradual Learning Algorithm (Boersma and Hayes, 2001).

b. *Rule-based* models: Only a small set of predicates (perhaps just one) participates in determining the output. Learning means testing, discarding, and modifying the current predicate set, and the model only does this to a small number of predicates at once. Example: RULEX (Nosofsky et al., 1994).

(8) Clear affinities to ideas in linguistics (here, phonology):

a. Is the *product of learning* (the grammar) a small list of rules that differs from language to language (Chomsky and Halle, 1968), or a weighing or ranking of a big universal constraint set (Prince and Smolensky, 1993)?

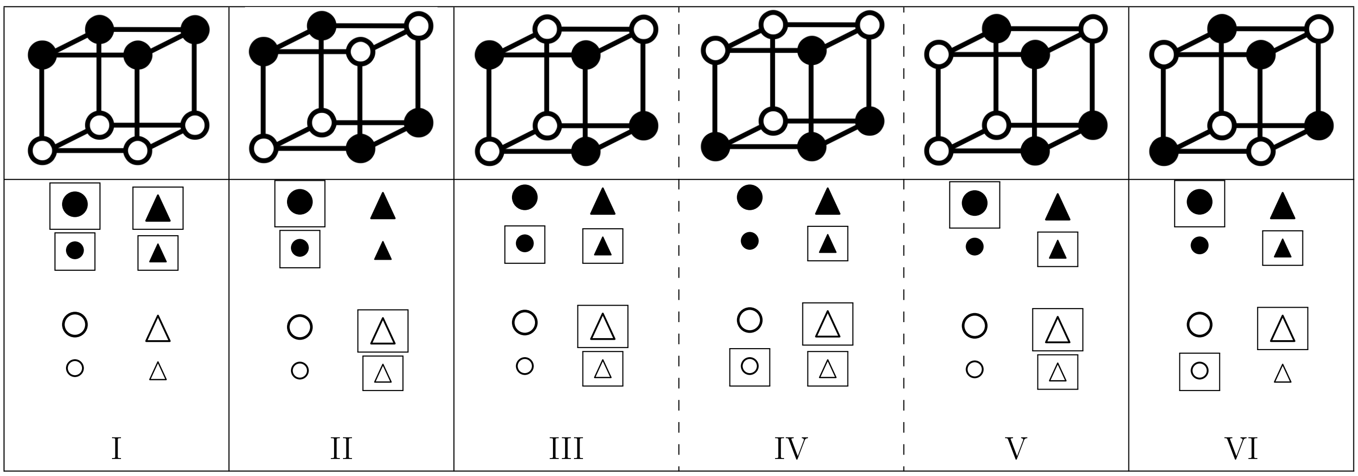
b. Is the *process of learning* the induction of rules by hypothesis generation and testing, or is the reranking or reweighting of constraints?

c. Do *biases in learning* emerge from the absence of (or reluctance to promote) particular predicates (Hayes and Wilson, 2008), or from preferences for syntactically simple hypotheses (Chomsky and Halle, 1968)?

(9) There are of course other kinds of pattern-learning model (for reviews, see Kruschke 2005; Ashby and Maddox 2005; Kruschke 2008), but this talk concentrates on the cue-based and rule-based architectural elements, which are used in many models in linguistics and psychology.

What experiments can we do to distinguish these two architectures?

(10) Research on structural inductive biases in non-linguistic pattern learning has focused on a family of pattern types first studied by Shepard et al. (1961):



a. Type I is a simple one-feature affirmation

b. Type II is IFF/XOR on two features

c. Types III–V need all three features, though subsets can be described with two

d. Type VI is a three-way IFF/XOR; every subset needs three features

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

(11) Architecture and structural inductive bias:

a. Rule-based models must test hypotheses in some order; usually, rules that are shorter or simpler by some standard are preferred (e.g., Nosofsky et al. 1994). ⇒ Generalizations that depend on fewer features are found faster ⇒ Favors Type II over Type IV.

b. Cue-based models are additive, so they learn patterns faster that are supported by multiple overlapping cues (Gluck and Bower, 1988; Pater et al., 2008; Pater and Moreton, 2012; Moreton et al., 2013). ⇒ "Family-resemblance" patterns are found especially fast ⇒ Favors Type IV over Type II.

A common interpretation, going back to Shepard et al. (1961), is that the $I > II > III, IV, V > VI$ order confirms that non-linguistic category learning is rule-based.

3 Implicit and explicit learning, and cue-based and rule-based architectures

(12) Some recent non-linguistic studies have found that the classic $I > II > III, IV, V > VI$ order is itself actually rather fragile, and that the advantage for Type II over Type IV can be reduced or even reversed by changing the stimuli or task conditions:

a. Training participants without right/wrong feedback (Love, 2002)

b. Not instructing participants to look for a rule (Love, 2002; Love and Markman, 2003; Lewandowsky, 2011; Kurtz et al., 2013)

c. Making stimulus dimensions harder to verbalize (Kurtz et al., 2013), or perceptually less-separable (Nosofsky and Palmeri, 1996)

(13) This has led to proposals in the psychology literature that there are two concurrent learning processes in non-linguistic pattern learning.

a. A cue-based process that is "implicit" (effortless, unconscious, gradual, does not need attention or working memory) and that learns Type IV (family resemblance) faster than Type II (IFF/XOR)

b. A cue-basedprocess that is "explicit" (effortful, conscious, abrupt, uses attention and working memory) and that learns Type II (IFF/XOR) faster than Type IV (family resemblance)

and that the conditions in (12) favor the use of one or the other (Ashby et al., 1998; Love, 2002; Maddox and Ashby, 2004; Smith et al., 2012). This proposal is one manifestation of a general paradigm in psychology (critically reviewed by Osman 2004; Evans 2008; Newell et al. 2011).

(14) Could the same be true for phonological learning?

a. Exp. 1: Are implicit and explicit processes both available? Do the same conditions favor one over the other as in non-linguistic pattern learning?

b. Exp. 2: Do the two processes differ in sensitivity to Type II (IFF/XOR) vs. Type IV (family-resemblance) structures, as in non-linguistic pattern learning?

4 Experiment 1: Implicit and explicit phonotactic learning

(15) Research questions: Are are explicit and implicit processes available in phonology, as in non-linguistic pattern learning? Do the same conditions favor their use in both domains?

(16) Conditions found in non-linguistic learning to favor explicit vs. implicit learning:

	Favors		
Condition	Explicit	Implicit	
Training	with feedback	no feedback	Love (2002)
Instructions	"seek a rule!"	don't mention rules	Love (2002); Love and Markman (2003); Lewandowsky (2011); Kurtz et al. (2013)
Features	verbalizable	not verbalizable	Nosofsky and Palmeri (1996); Kurtz et al. (2013)

(17) Diagnostic symptoms of explicit vs. implicit learning:

Symptom	Characteristic of Explicit	Characteristic of Implicit	
Report rule seeking/finding/use	yes	no	
Can state correct rule	yes	no	
Shape of learning curve	abrupt	gradual	Smith et al. (2004)
Distribution of test-phase performance	bimodal	unimodal	Kurtz et al. (2013)
Structural bias	$II > IV$	$II < IV$	Love (2002); Kurtz et al. (2013)

(18) Basic idea: Using simple Type I patterns, vary the conditions in (16) and see if they change the signatures in (17).

4.1 Methods

(19) Stimuli:

a. Audio: Vowel-initial nonwords, two or three syllables, initial or second-syllable stress: [(əC)VCəC] and [(Cə)CəCVC. $C \in$ {p b t d f v s z}; $V \in$ {i ɪ e ɛ u ʊ o ɔ}. Recorded by male native speaker of American English from Upper Midwest.

Consonants

	Lab		Cor	
voiced	−	+	−	+
−cont	p	b	t	d
+cont	f	v	s	z

Stressed vowels

	−back		+back	
tense	+	−	+	−
+high	i	ɪ	u	ʊ
−high	e	ɛ	o	ɔ

Prosodic shapes

	Disyllabic	Trisyllabic
$\acute{\sigma}_1$	VCəC	VCəCəC
$\acute{\sigma}_2$	əCVC	əCVCəC

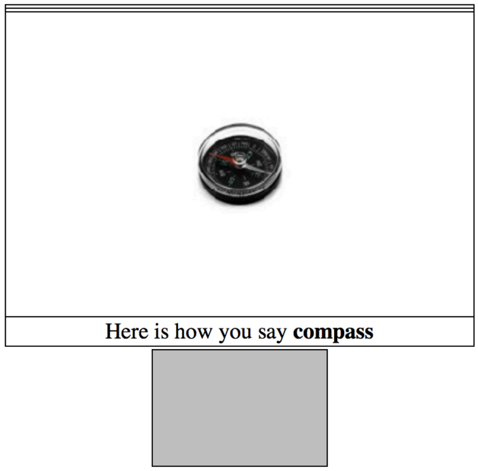
b. Visual: 160 images of objects, 20 in each of the cells defined by edible/inedible × large/small × long/compact.

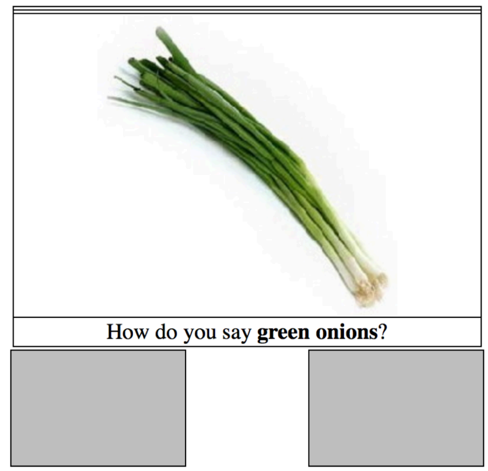
(20) Phonological pattern (Type I): For each participant, the stimuli were divided into "legal" and "illegal" classes on the basis of *one* of the following properties:

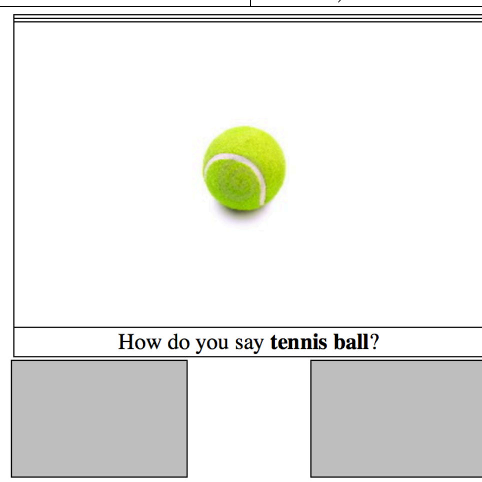
3	two syllables	three syllables
4	initial stress	second-syllable stress
5	all consonants identical	otherwise
6	stressed V is back	stressed V is front
7	all Cs are stops	all Cs are fricatives
8	all Cs are labial	all Cs are coronal

randomizing which value was positive. Each participant got a unique instantiation of their pattern.

(21) Procedure: Participants were recruited for a study on learning words in an artificial language using Amazon Mechanical Turk (Sprouse, 2011).

<table>
<tr><td>PHASE</td><td colspan="2">EVENTS</td></tr>
<tr><td>Initialization</td><td colspan="2">Welcome; sound check</td></tr>
<tr><td rowspan="2">Instructions</td><td>No-Feedback</td><td>Feedback</td></tr>
<tr><td>Learn words</td><td>Learn to tell correct word from foil; look for rule that will let you get it 100% right</td></tr>
<tr><td>Training</td><td>

See picture, hear correct (=legal) word. 4 repetitions of 32 pictures and words.</td><td>

See picture, hear correct word and foil (= legal and illegal); choose one; hear right/wrong feedback. 4 repetitions of 32 pictures and word/foil pairs; stopped early if two consecutive perfect blocks of 8 (“met criterion”).</td></tr>
<tr><td>Test</td><td colspan="2">

See picture, hear correct word and foil; choose one; no feedback. 32 word/foil pairs and pictures, all new.</td></tr>
<tr><td>Debriefing</td><td colspan="2">Questionnaire about learning strategy, demographic information</td></tr>
</table>

4.2 Results

4.2.1 Did the Feedback condition facilitate reported rule learning? *Yes.*

(22) Participants in the Feedback condition were indeed significantly more likely to report "rule use" (seeking, finding, or stating a rule) than those in the No-Feedback condition ($\chi^2 = 11.1606$ on one d.f., $p = 0.0008355$).

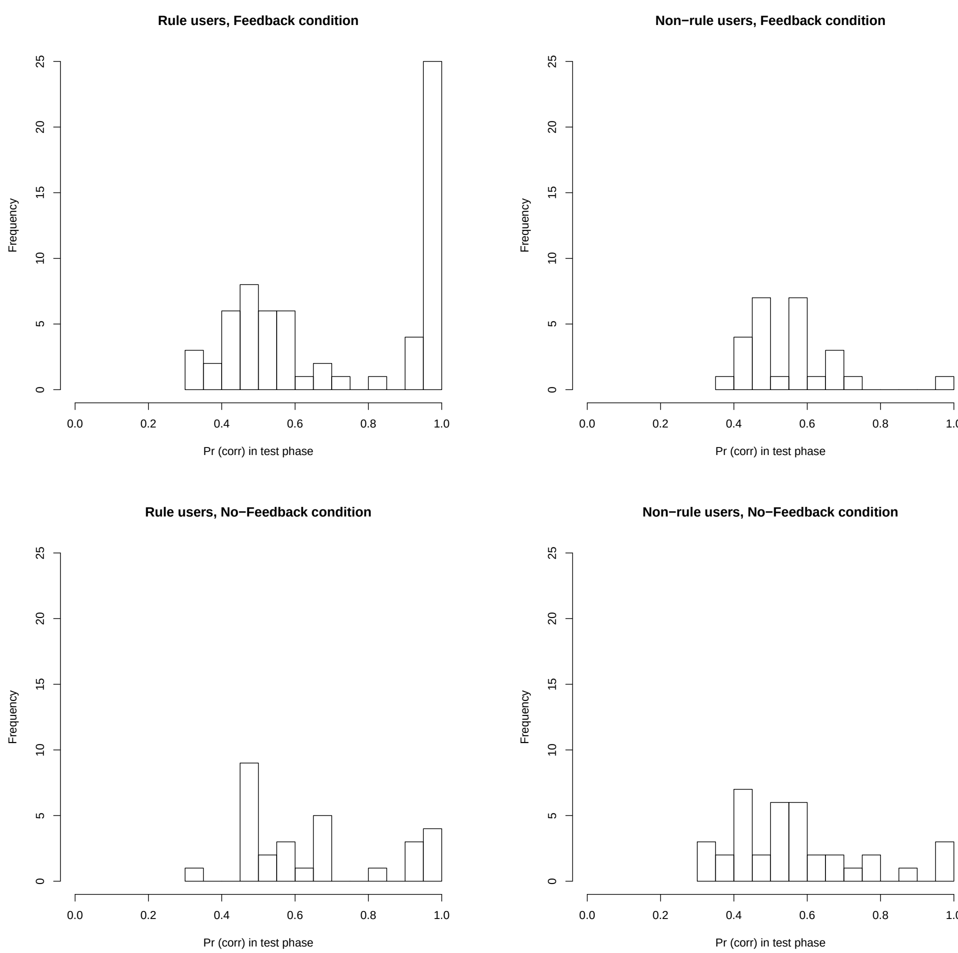
Rule use	No-Feedback	Feedback
FALSE	37	26
TRUE	28	64

(23) The correct property was identified significantly more often in the Feedback than the No-Feedback condition ($\chi^2 = 5.6638$ on one d.f., $p = 0.01732$):

Correct property	No-Feedback	Feedback
FALSE	63	74
TRUE	3	17

4.2.2 Was rule-users' performance more bimodal? *Yes.*

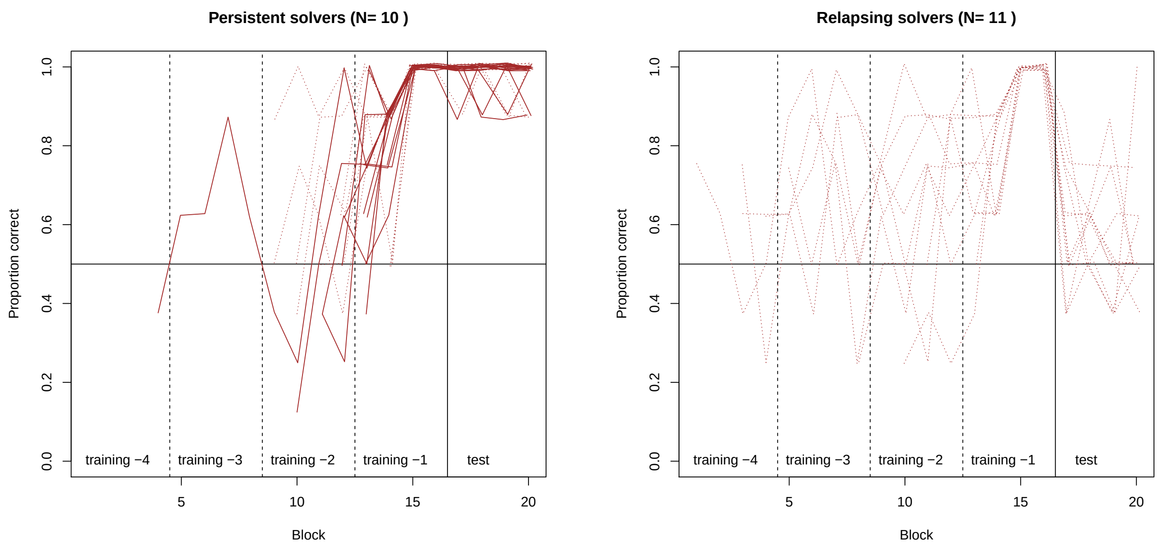
(24) The test-phase performance of those who reported rule use was bimodally distributed (either near-perfect, or near-chance). The performance of those who did not report rule use was unimodally distributed near chance.



4.2.3 Did explicit rule learners learn more abruptly? *Yes.*

(25) These points apply to the Feedback condition only, because the No-Feedback condition does not provide the necessary data (learning curves).

(26) *Persistent vs. relapsing solvers*: In the Feedback condition, most participants who reached criterion in the training phase (16 consecutive correct responses) did very well on the test (at least 80% correct). Some, however, relapsed to near-chance performance. The following individual block-by-block learning curves are aligned to the beginning of the criterion run:



Solid lines show participants who named the correct property. Solvers who did so were significantly less likely to relapse in the test phase ($\chi^2 \approx$ 10.1153 on one d.f., $p \approx$ 0.00147):

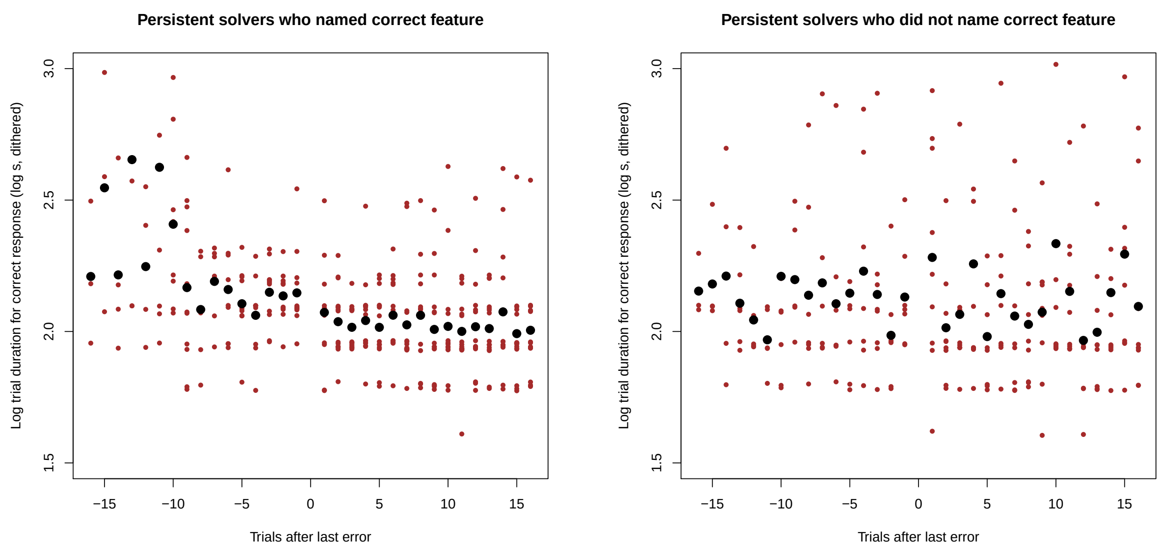
Correct property	Relapsing	Persistent
FALSE	11	10
TRUE	0	17

Relapsing solvers had probably not learned the pattern, but rather had simply memorized the individual training items.

(27) *Abruptness of training-phase improvement*: We expected that among persistent solvers (i.e., people who had learned the pattern, not just the stimuli), those who stated the correct property (i.e., people who definitely had found a correct explicit rule) would show abrupt improvement at criterion, while those who did not state the correct property would not. An analysis of the last 16 trials preceding the last error found only a marginal trend in that direction:

```
 Formula: correct ~ corrfeat + (1 | subjid)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.5341     0.2807   5.465 4.62e-08 ***
corrfeatTRUE  -0.5818     0.3457  -1.683   0.0923 .
```

(28) *Acceleration in training-phase response latency*: We hypothesized that participants would respond slower when they were searching for a rule, and faster once they had found one. To test this hypothesis, we compared individual trial durations before and after the last error. Solvers who state the correct property accelerate more after their last error than solvers who do not state the correct property. (Big dots are means; little dots are individual responses.)




```
Formula: log(trial_duration) ~ corrfeat * I(wrt.last.err > 0) + (1 | subjid)
                                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)                      2.23653    0.05246  29.60000  42.631  < 2e-16 ***
corrfeatTRUE                     0.08511    0.06688  30.40000   1.273 0.212837
I(wrt.last.err > 0)TRUE         -0.11793    0.03681 669.60000  -3.204 0.001420 **
corrfeatTRUE:I(wrt.last.err > 0)TRUE -0.17519    0.04722 671.80000  -3.710 0.000224 ***
```

4.2.4 Summary: Exp. 1

(29) The *Feedback* condition elicited traits of explicit and learning, while the *No-Feedback* condition elicited traits of implicit learning:

	Condition	
Signature	*Feedback*	*No-Feedback*
Report rule seeking/finding/use	more	less
Can state correct rule	more	less
Shape of learning curve	abrupt (?)	(not tested)
Distribution of test-phase performance	bimodal	unimodal
Structural bias	(not tested)	(not tested)

⇒ In phonotactic learning, as in non-linguistic pattern learning, implicit and explicit systems are available, and they are facilitated or inhibited by similar factors.

5 Experiment 2: Cue-based and rule-based phonological learning

(30) Research question: Do the implicit and explicit systems found in Exp. 1 differ in their receptiveness to Type II vs. Type IV patterns, as they do in non-linguistic learning?

5.1 Methods

(31) Of the six phonological properties in (20), we chose the three that had elicited the best test-phase performance in Exp. 1 (two vs. three syllables, initial vs. second-syllable stress, and fricatives vs. stops) to serve as the three cube axes for all participants in Exp. 2.

Conditions were (Type I, Type II, Type IV) × (*No-Feedback*, *Feedback*), with approximately 48 participants in each of the 6 cells.

(32) The participant pool and procedure were like those in Exp. 1, except that the questionnaire included check boxes for some of the main strategies that participants had reported in Exp. 1. In particular,

a. Participants in Exp. 2 were asked separately whether they had *sought* a rule in the training phase, and whether they had *used* one in the test phase.

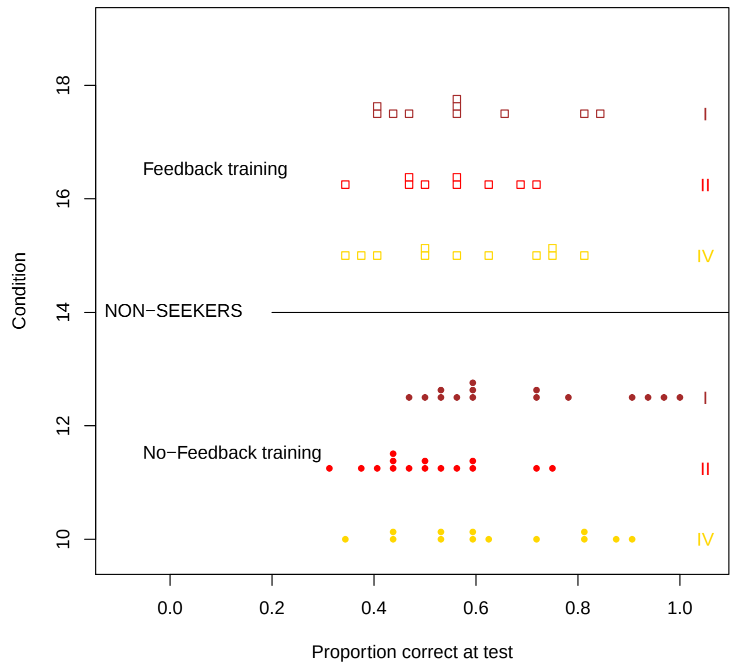
b. They were also asked whether, in the test phase, they had deliberately chosen foils that were maximally *unlike* what they had heard in the training phase. Data from those who said yes was excluded.

5.2 Results

(33) The *Feedback* vs. *No-Feedback* manipulation had similar effects to Exp. 1 on reports of rule-seeking.

5.2.1 Do implicit learners perform worse on Type II than Types I and IV? *Yes.*

(34) If non-rule-seekers were using an implicit cue-based process, their test-phase performance should decline in the order $I > IV > II$ (Love, 2002; Kurtz et al., 2013). This figure shows the test-phase proportion correct for each non-rule-seeker.

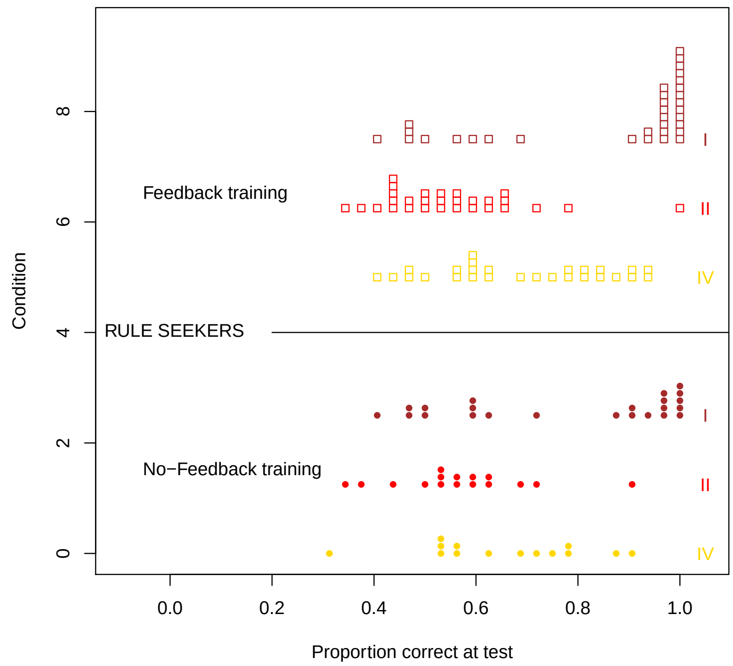


(35) For non-rule-seekers who were trained without feedback, the probability of a pattern-conforming response was significantly above chance for Type I, and significantly below Type I (near chance) for Type II. Type IV was not significantly below Type I. Training with feedback significantly reduced Type I performance, but didn't affect Type II performance much.

```
Formula: correct ~ Type * TrainingGroup + (1 | subjid)
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                0.9308     0.1906   4.884 1.04e-06 ***
TypeII                    -0.8935     0.2638  -3.387 0.000707 ***
TypeIV                    -0.3294     0.2755  -1.196 0.231710
TrainingGroupTwoAFC       -0.6119     0.2945  -2.078 0.037749 *
TypeII:TrainingGroupTwoAFC  0.7797     0.4191   1.860 0.062824 .
TypeIV:TrainingGroupTwoAFC  0.3467     0.4153   0.835 0.403771
```

5.2.2 Do explicit learners perform better on Type II than Type IV? *No.*

(36) Participants who reported rule-seeking were presumably using an explicit, rule-based strategy, which in non-linguistic experiments makes Type II patterns easier relative to Type IV. However, performance in the Type II condition was at chance:



What happened?

(37) *Explicit rule learning failed on Types II and IV* As in Exp. 1, participants often reported the correct property for Type I. In the Type II and IV conditions, however, where the rule depended on two or three properties, it was rare for a participant to report even one:

		Properties stated			
Training group	Type	0	1	2	3
No-Feedback	I	30	8	–	–
	II	24	4	3	–
	IV	25	2	0	0
Feedback	I	22	21	–	–
	II	34	4	0	–
	IV	26	8	4	0

a. *Type II*: Only *one* participant stated the rule approximately correctly:"It seemed like the shorter words ended with b or d [or p or t — EM]. The longer words with 3+ syllables seemed to end with f or v [or s or z — EM].". (They got 91% right on the test.)

b. *Type IV*: No one named all three relevant properties. However, by identifying one property, it was possible to get 75% correct using an explicit Type I rule, and in fact the 14 Type IV participants who did name at least one property got, on average, 76% correct on the test.

c. Statistically, test-phase performance was marginally *better* on Type IV than Type II, in both the *No-Feedback* and the *Feedback* conditions.

5.2.3 Summary: Exp. 2

(38) Participants who reported learning implicitly did indeed perform better on Type IV than Type II, and the Type IV advantage was weakened for explicit learners.

However, the implicit/explicit difference was not due to the better Type II performance of explicit learners, but to their worse Type IV performance.

(39) ⇒ Even with highly separable, highly verbalizable dimensions,

a. *implicit* phonotactic learning shows the $IV > II$ bias characteristic of cue-based learning;

   (i) Like non-linguistic pattern learning (Love, 2002; Kurtz et al., 2013)

   (ii) Like phonotactic learning with hard-to-verbalize, hard-to-separate features (Moreton and Pertsova, 2014; Moreton et al., 2013) (unless they are agreement/disagreement patterns (Moreton, 2008, 2012)).

b. *explicit* phonotactic learning is quite difficult for patterns involving more than one dimension. This is *unlike* non-linguistic pattern learning; (Shepard et al., 1961; Haygood and Bourne, 1965; Nosofsky et al., 1994; Love, 2002; Kurtz et al., 2013, e.g.,)

6 Discussion

(40) *Summary*: Research on non-linguistic pattern learning has found evidence of distinct implicit and explicit learning processes, which have different computational architectures, are facilitated by different conditions, and solve different kinds of inductive problem most efficiently. This talk has presented new evidence that

a. Implicit and explicit processes are available for phonotactic learning of simple (Type I) patterns (Exp. 1);

b. The implicit process is more successful with Type IV (family-resemblance) than Type II (IFF/XOR) patterns, consistent with a cue-based architecture;

c. The explicit process (surprisingly) has great difficulty with Type II, unlike what is found in non-linguistic learning.

This part of the talk addresses some questions and opportunities that arise out of these findings.

(41) Implications for interpreting phonological learning experiments: There are two processes, which are elicited by different conditions, and which may have different inductive biases. Differences in experimental outcomes may be due to differences in which process is engaged by the experimental conditions, or to within-condition differences between participants as to which process they favored.

(42) The search for *substantive* inductive bias (bias that cares about the real-world content of the features, not just their logical arrangement) has yielded contradictory results.[1] Is that because the two processes differ in their sensitivity to phonetic substance, and different experiments (or participants) favor different processes?

(43) What is the connection between the implicit and explicit processes observed in the lab, and real L1 or L2 acquisition? Are there other learning processes besides the two discussed here? For example,

[1]See review in Moreton and Pater 2012a,b. Since then there have been new findings regarding the structural properties of locality (Finley, 2011, 2012; McMullin, 2013). Another set of new results concerning "saltation" in UR→SR mappings is open to either structural or substantive interpretation (White, 2014; White and Sundara, 2014).

a. “aha!”-type insight in problem solving, which is abrupt but unconscious (Metcalfe and Wiebe, 1987; Bowden et al., 2005; Žauhar et al., 2014)

b. “Across-the-board” saltations in L1 acquisition (Smith, 1973; Vihman and Velleman, 1989; Levelt and van Oostendorp, 2007; Gerlach, 2010; Guy, 2013). Does “By Jove, I think she’s got it!” really happen?

(44) Explicit learning is often regarded as an undesirable contamination; we don’t want our participants to “use a strategy” or “solve crossword puzzles”, because those processes are hypothesized to be remote from natural language acquisition (and hence to have no influence on typology).

However, a lot of natural language acquisition is second-language acquisition, and implicit vs. explicit is a major theme in the L2 morphosyntax literature (see recent review in Lichtman 2012). Perhaps explicit learning is worth more serious theoretical attention.

(45) Does natural-language typology show the imprint of a cue-based or a rule-based learning process? If the “phonologically active classes” in P-Base (**?**) are assigned to SHJ Types on the basis of the *SPE* features, and compared to an equal number of size- and inventory-matched random classes, the real classes exceed the random ones in the same $I > II > III, IV, V > VI$ order (Moreton and Pertsova, 2014):

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

That seems to point to a rule-based system, since Type II (IFF/XOR) outnumbers Type IV (family resemblance). How can we reconcile this with the robustness of the Type IV > Type II advantage in implicit phonotactic learning? (Do the phonetic precursors look more like Type II than Type IV, for instance?)

(46) The literature on structural biases in non-linguistic learning by humans is immense:

a. Two-feature relations: AND > OR > IFF/XOR (Bruner et al. 1956, Ch. 6, Neisser and Weene 1962; Hunt and Kreuter 1962; Conant and Trabasso 1964; Haygood and Bourne 1965; King 1966; Snow and Rabinovitch 1969; Gottwald 1971a,b; Lee 1981 . . .

b. . . . but inter-dimensional IFF/XOR are much easier (review and phonological analogues in Moreton 2012)

c. Effect of response labels: If the response is A vs. B, then AND and OR are the same pattern. If it’s A vs. not-A, then AND gets easier and OR gets harder (Gottwald 1971b; Peters and Denny 1971; morphological analogue in Pertsova 2012)

d. Base-rate neglect: When trained on a probabilistic non-linguistic pattern, participants overestimate the degree to which a feature that is characteristic of a rare category is predictive of it (Gluck and Bower, 1988; Nosofsky et al., 1992; Kahneman and Tversky, 1996). ⇒ Generalization of “minority” patterns if they have sufficiently characteristic phonology?

(Schaffhausen German: /o/-lowering triggered by adjacent /r m n ŋ/ in the city, /r t ts s z ʃ ʒ/ in some nearby villages; thought by Cristiá et al. (2013) to be generalizations from original /r/. Is generalization from a rare to a frequent class more common than the reverse?)

e. Many, many more (see, e.g., Kahneman 2011).

(47) Do these biases affect linguistic pattern learning too? And if so,

a. Do they appear in L2 learning only, or L1 as well?

b. Do they appear only in the early stages of learning, or are they persistent throughout?

c. Do any of them shape typology?

References

Ashby, F. G., L. A. Alfonso-Reese, A. U. Turken, and E. M. Waldron (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review 105*(3), 442–481.

Ashby, F. G. and W. T. Maddox (2005). Human category learning. *Annual Review of Psychology 56*, 149–178.

Bell, A. (1970). *A state-process approach to syllabicity and syllabic structure.* Ph. D. thesis, Stanford University.

Bell, A. (1971). Some patterns of the occurrence and formation of syllabic structure. *Working Papers on Language Universals 6*, 23–138.

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

Bowden, E. M., M. Jung-Beeman, J. Fleck, and J. Kounious (2005). New approaches to demystifying insight. *TRENDS in Cognitive Sciences 9*(7), 322–328.

Bruner, J. S., J. J. Goodnow, and G. A. Austin (1956). *A study of thinking.* New York: John Wiley and Sons.

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

Conant, M. B. and T. Trabasso (1964). Conjunctive and disjunctive concept formation under equal-information conditions. *Journal of Experimental Psychology 67*(3), 250–255.

Cristiá, A., J. Mielke, R. Daland, and S. Peperkamp (2013). Similarity in the generalization of implicitly learned sound patterns. *Laboratory Phonology 4*, 259–285.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology 49*, 255–278.

Finley, S. (2011). The privileged status of locality in consonant harmony. *Journal of Memory and Language 65*(74–83).

Finley, S. (2012). Testing the limits of long-distance learning: Learning beyond a three-segment window. *Cognitive Science 36*(4), 740–756.

Gallistel, C. R., A. L. Brown, S. Carey, R. Gelman, and F. C. Keil (1991). Lessons from animal learning for the study of cognitive development. In S. G. Carey and R. G. Gelman (Eds.), *The epigenesis of mind*, Chapter 1, pp. 3–36. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gerlach, S. R. (2010). *The acquisition of consonant feature sequences: harmony, metathesis, and deletion patterns in phonological development.* Ph. D. thesis, University of Minnesota.

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

Gordon, M. (2004). Syllable weight. In B. Hayes, R. Kirchner, and D. Steriade (Eds.), *Phonetically-based phonology*, pp. 277–312. Cambridge, England: Cambridge University Press.

Gottwald, R. L. (1971a). Attribute-response correlations in concept attainment. *American Journal of Psychology 84*(3), 425–436.

Gottwald, R. L. (1971b). Effects of response labels in concept attainment. *Journal of Experimental Psychology 91*(1), 30–33.

Greenberg, J. H. (1978). Diachrony, synchrony, and language universals. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Eds.), *Universals of human language, volume 1, method and theory*, pp. 61–91. Stanford, California: Stanford University Press.

Guy, G. R. (2013). Linking usage and grammar: generative phonology, exemplar theory, and variable rules. *Lingua xx*(xx), xx–xx. (to appear).

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

Haygood, R. C. and L. E. Bourne (1965). Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review 72*(3), 175–195.

Heinz, J. and W. Idsardi (2011). Sentence and word complexity. *Science 333*(6040), 295–297.

Hunt, E. B. and J. M. Kreuter (1962, December). The development of decision trees in concept learning: III, learning the logical connectives. Working Paper 92, University of California, Los Angeles, Western Management Science Institute.

Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic studies offered to Joseph Greenberg: second volume: phonology*, pp. 407–418. Saratoga, California: Anma Libri.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Strauss, and Giraux.

Kahneman, D. and A. Tversky (1996). On the reality of cognitive illusions. *Psychological Review 103*(3), 582–591.

King, W. L. (1966). Learning and utilization of conjunctive and disjunctive classification rules: a developmental study. *Journal of Experimental Child Psychology 4*(3), 217–231.

Kruschke, J. K. (2005). Category learning. In K. Lamberts and R. L. Goldstone (Eds.), *The handbook of cognition*, Chapter 7, pp. 183–201. London: Sage.

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge handbook of computational psychology*, Chapter 9, pp. 267–301. New York: Cambridge University Press.

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.

Lee, S.-S. (1981). The role of memory aids in two instructiona paradigms: learning and utilization of logical rules. *Instructional Science 10*, 123–133.

Levelt, C. and M. van Oostendorp (2007). Feature co-occurrence constraints in L1 acquisition. *Linguistics in the Netherlands 24*(1), 162–172.

Lewandowsky, S. (2011). Working memory capacity and categorization: individual differences and modelling. *Journal of Experimental Psychology: Learning, Memory, and Cognition 37*(3), 720–738.

Lichtman, K. M. (2012). *Child-adult differences in implicit and explicit second language learning.* Ph. D. thesis, University of Illinois at Urbana-Champaign.

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. and A. B. Markman (2003). The nonindependence of stimulus properties in human category learning. *Memory and Cognition 31*(5), 790–799.

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

Maddox, W. T. and F. G. Ashby (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes 66*, 309–332.

McCarthy, J. J. (1988). Feature Geometry and dependency: a review. *Phonetica 45*, 84–108.

McMullin, K. (2013). Learning consonant harmony in artificial languages. In *Actes du congrès annuel de l'Association candienne de linguistique 2013/Proceedings of the 2013 Annual Conference of the Canadian Linguistic Association*, pp. xx–xx.

Metcalfe, J. and D. Wiebe (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition 15*(3), 238–246.

Mitchell, T. M. (1980 [1990]). The need for biases in learning generalizations. In J. W. Shavlik and T. G. Detterich (Eds.), *Readings in machine learning*, Chapter 2.4.1, pp. 184–191. San Mateo, California: Morgan Kaufman.

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.

Moreton, E., J. Pater, and K. Pertsova (2013, October). Phonological concept learning. MS submitted to *Cognition*.

Moreton, E. and K. Pertsova (2014). Pastry phonotactics: is phonological learning special? In H.-L. Huang, E. Poole, and A. Rysling (Eds.), *Proceedings of the 43rd Annual Meeting of the Northeast Linguistic Society, City University of New York*. Amherst, Massachusetts: Graduate Linguistics Students' Association.

Neisser, U. and P. Weene (1962). Hierarchies in concept attainment. *Journal of Experimental Psychology 64*(6), 640–645.

Newell, B. R., J. C. Dunn, and M. Kalish (2011). Systems of category learning: fact or fantasy? In B. Ross (Ed.), *The psychology of learning and motivation*, Volume 54, Chapter 6, pp. 167–215. Academic Press.

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., J. K. Kruschke, and S. C. McKinley (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition 18*(2), 211–233.

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.

Ohala, J. J. (1992). What's cognitive, what's not, in sound change. In G. Kellerman and M. D. Morissey (Eds.), *Diachrony within synchrony: language history and cognition*, pp. 309–355. Frankfurt am Main: Peter Lang.

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

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin and Review 11*(6), 988–1010.

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.

Pater, J., E. Moreton, and M. Becker (2008, November). Simplicity biases in structured statistical learning. Poster presented at the Boston University Conference on Language Development.

Pertsova, K. (2012). Logical complexity in morphological learning. To appear in *Proceedings of the Berkeley Linguistics Society*.

Peters, K. G. and J. P. Denny (1971). Labelling and memory effects on categorizing and hypothesizing behavior for biconditional and conditional conceptual rules. *Journal of Experimental Psychology 87*(2), 229–233.

Pinker, S. (1979). Formal models of language learning. *Cognition 7*, 217–283.

Prince, A. and P. Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Department of Linguistics, Rutgers University.

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., M. E. Berg, R. G. Cook, M. S. Murphy, M. J. Crossley, J. Boomer, B. Spiering, M. J. Beran, B. A. Church, F. G. Ashby, and R. C. Grace (2012). Implicit and explicit categorization: a tale of four species. *Neuroscience and Biobehavioral Reviews 36*(10), 2355–2369.

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.

Smith, N. (1973). *The acquisition of phonology: a case study*. Cambridge, England: Cambridge University

Press.

Snow, C. E. and M. S. Rabinovitch (1969). Conjunctive and disjunctive thinking in children. *Journal of Experimental Child Psychology 7*(1), 1–9.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgements in linguistic theory. *Behavior Research Methods 43*(1), 155–167.

Vihman, M. M. and S. Velleman (1989). Phonological reorganization: a case study. *Language and Speech 32*, 149–170.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition 130*, 96–115.

White, J. and M. Sundara (2014). Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition 133*(1), 85–99.

Žauhar, V., I. Bajšanski, and D. Domijan (2014). Metacognitive monitoring of rule-based category learning tasks. In P. Bernardis, C. Fantoni, and W. Gerbin (Eds.), *Proceedings of the Trieste Symposium on Perception and Cognition, November 27th–28th, 2014*, pp. 162–165.