BUCLD 40 Proceedings
To be published in 2016 by Cascadilla Press
Rights forms signed by all authors

Implicit and explicit processes in phonotactic learning

Elliott Moreton and Katya Pertsova*

Recent years have seen a proliferation of adult phonological-learning studies ("artificial-language" experiments) employing a wide array of experimental tasks, instructions, and materials (reviewed in Moreton & Pater 2012a,b), in the hope of gaining experimental access to the inductive processes underlying first- or second-language acquisition. But there has been little investigation into what is actually going on in these experiments. Do different experimental situations engage different learning processes? If so, do those processes have different inductive biases? How are they related to the processes involved in L1 and L2 acquisition? Answers to these questions have implications for both methodology in particular, and cognitive science in general.

Studies of non-linguistic (mainly visual) pattern learning have led psychologists to hypothesize two concurrent learning processes that have different properties and that are facilitated by different experimental conditions (Ashby et al., 1998; Love, 2002; Maddox & Ashby, 2004; Smith et al., 2012). Here we will call them the *explicit system* and the *implicit system*. The explicit system is effortful, conscious, abrupt, and rule-based (e.g., it can be modelled as serial testing of featurally-simple hypotheses); it demands focused attention and working memory, and its use is facilitated by training with right/wrong feedback, instructions to seek a rule, and the use of easily verbalizable stimulus features. The implicit system is effortless, unconscious, gradual, and cue-based (i.e., it can be modelled as weight update on an array of property detectors); it does not need attention or working memory, and its use is facilitated by training without feedback, instructions that do not mention rules, and non-verbalizable stimulus features. This proposal is one manifestation of a more general idea in psychology (reviewed by Osman 2004; Evans 2008; Newell et al. 2011). The two systems also differ in sensitivity to different pattern structures (see below).

This paper presents two experiments. Experiment 1 asks whether implicit and explicit processes are available for phonological learning, and, if so, whether they are facilitated by the same conditions as in non-linguistic pattern learning. Experiment 2 asks whether the two processes differ in sensitivity to different pattern types in the same way as in non-linguistic learning. This study contributes to a larger program, the comparative study of inductive learning across domains (Pertsova, 2012; Pater & Moreton, 2012; Moreton, 2012; Moreton et al., in press).

*Department of Linguistics, University of North Carolina, Chapel Hill, NC 27510–3155 U.S.A. Email: (moreton|pertsova)@unc.edu.

1 Experiment 1: Implicit and explicit learning of single-feature patterns

Indicators of explicit rather than implicit learning include: self-report of rule-seeking, rule-finding, or rule-use (Bruner et al., 1956; Ciborowski & Cole, 1972); ability to state the correct rule (Ciborowski & Cole, 1973); an abrupt jump in the learning curve (Smith et al., 2004) and abrupt acceleration in response times (Haider & Rose, 2007) when the solution is found; and a bi-modal distribution of performance between "solvers" who are close to perfect, and "non-solvers" who are close to chance (Love, 2002; Kurtz et al., 2013).

The experimental conditions which facilitate explicit over implicit learing in non-linguistic experiments include training with feedback (Love, 2002), instructions to seek a rule (Love, 2002; Lewandowsky, 2011; Kurtz et al., 2013), and the use of perceptually-separable, easily-verbalizable features (Nosofsky & Palmeri, 1996; Kurtz et al., 2013). The strategy of Experiment 1 is to manipulate these conditions in a phonotactic-learning experiment, and see if they have the same effect on the indicators as they have in non-linguistic experiments.

1.1 Stimuli, methods, and participants

Participants were recruited for a study on learning grammatical gender in an artificial language. The experiment was run over the World Wide Web via the Amazon Mechanical Turk platform (Sprouse, 2011). Participants were required to have completed at least 1000 previous Mechanical Turk assignments, with an approval rate of at least 95%. Each stimulus *Word* consisted of a picture paired with an audio nonword of English (recorded by a native speaker of American English). The nonwords were vowel-initial, di- or trisyllabic, with stress on the first or second syllable: $\{\acute{V}C\text{ə}C,\ \acute{V}C\text{ə}C,\ \text{ə}C\acute{V}C,\ \text{ə}C\acute{V}C\}$. Consonants were voiced or voiceless, labial or coronal, fricatives or stops: {p b t d f v s z}. Stressed (non-schwa) vowels were front or back, high or low, tense or lax: {i ɪ e ɛ u ʊ o ɔ}. Pictures were 160 images of familiar objects, 20 in each of the cells defined by the features edible/inedible, long (I-shaped)/compact (O-shaped), large (foot-sized)/small (finger-sized).

For each participant, the pictures were randomly paired with audio nonwords to make Words. Then one of the following nine phonological or semantic features was chosen: edible/inedible, long/compact, large/small, disyllabic/trisyllabic, first/second-syllable stress, all consonants identical/some consonants different, stressed vowel is back/front, all consonants are stops/fricatives, all consonants are labial/coronal. The Words were divided into positive and negative categories on the basis of that feature, and the positive and negative categories were randomly assigned to be "feminine" or "masculine". The experimental procedure is illustrated in Figure 1.

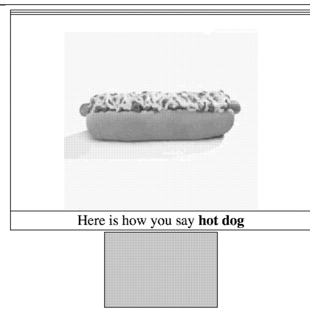
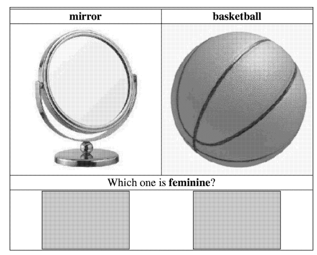
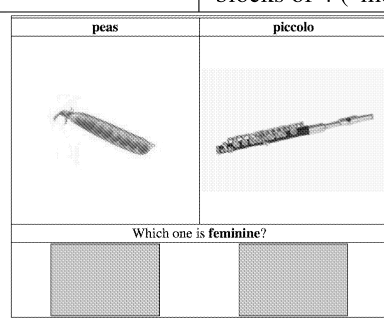
PHASE	EVENTS	
Initialization	Welcome; sound check	
Instructions	*No-Feedback*	*Feedback*
	Learn to recognize feminine-(masculine-)gender words in an artificial language	Learn to tell feminine-gender words from masculine-gender ones; look for rule that will let you get it 100% right
Training	 See picture, hear word; all are feminine (masculine). 4 repetitions of 32 Words.	 See and hear two Words, one feminine, one masculine. Choose the one you think is feminine (masculine); hear right/wrong feedback. 32 f. Words and 32 m. Words, recombined from trial to trial; up to 128 trials. Stopped early if 4 consecutive perfect blocks of 4 ("met criterion").
Test	 See picture, hear correct word and foil; choose one; no feedback. 32 trials, all new stimuli.	
Debriefing	Questionnaire about learning strategy	

Figure 1. Procedure of Experiments 1 and 2.

One part of the questionnaire asked: "How did you approach the learning task (the first part of the experiment)? Please choose all that apply: ☐ Went by intuition or gut feeling. ☐ Tried to memorize the words. ☐ Tried to find a rule or pattern. Please describe what you did in as much detail as you can. If you looked for a rule, what rules did you try?" (followed by a text box for a free-form response). A subsequent question asked, "How did you approach the test (the second part of the experiment)? Please choose all that apply: ☐ Chose words that sounded *similar* to the words I'd studied. ☐ Chose words that sounded *different* from the words I'd studied. ☐ Chose words that fit a rule or pattern. Again, please describe what you did in as much detail as you can. If you used a rule, what was it?" (followed again by a text box).

A total of 211 participants completed the experiment. Of these, 20 were excluded from analysis (5 reported a non-English L1, 7 reported taking written notes, 6 reported choosing test-phase responses that were maximally *un*like what they were trained on, 2 fell below the minimum performance criterion of at least 10 correct answers in the test phase), leaving 191 valid participants (98 *No-Feedback* and 93 *Feedback*).

1.2 Results

For each of the 211 participants, in random order, their responses to the open-ended training-phase and test-phase strategy questions were displayed to the experimenter simultaneously, together with a statement of the correct rule (e.g., "masculine $\Leftrightarrow$ disyllabic"), but with no information about whether the participant was in the *Feedback* or *No-Feedback* condition. The first author scored these responses *together* as true or false for the following: (1) Did they state a rule, i.e., a connection between their responses and some public property of the stimulus. (2) Did they state the correct rule, i.e., one that would allow a correct decision in every case? (3) Did they state a rule that was partly correct, i.e., it would allow better than chance performance, assuming that performance was at chance on any cases not addressed by the rule? Scoring was repeated three weeks later in a different random order. Agreement between the two scoring sessions was 93.5% for (1), 95.9% for (2), and 99.5% for (3).

Effect of experimental condition on learning strategy: Participants in the *Feedback* condition were significantly more likely than those in the *No-Feedback* condition to report seeking a rule in training by checking the "Tried to find a rule or pattern" box, and to report using a rule in the test phase by checking the "Chose words that fit a rule or pattern" box (Table 1).

Hence the *No-Feedback*/*Feedback* manipulation affected self-reported rule seeking and rule use. However most people reported rule-seeking, and nearly half reported rule-use, even in the *No-Feedback* condition where they were not explicitly instructed to do so. *No-Feedback* participants were significantly more likely than *Feedback* participants to report memorizing individual stimuli (which recurred in the training phase, but not in the test phase, see Figure 1), but there was no significant difference in self-report of using intuition (Table 2).

Table 1. Report of rule-seeking in training phase, and rule-use in test phase, as a function of training group, Experiment 1.

	Sought rule	
Training group	FALSE	TRUE
No-Feedback	42	56
Feedback	17	76

χ^2 = 12.375, df = 1, p = 0.0004351

	Reported rule use	
Training group	FALSE	TRUE
No-Feedback	56	42
Feedback	36	57

χ^2 = 5.7768, df = 1, p = 0.01624

Table 2. Report of stimulus memorization and use of intuition as a function of training group, Experiment 1.

	Memorized	
Training group	FALSE	TRUE
No-Feedback	53	45
Feedback	71	22

χ^2 = 9.4301, df = 1, p = 0.002135

	Used intuition	
Training group	FALSE	TRUE
No-Feedback	67	31
Feedback	54	39

χ^2 = 1.7604, df = 1, p = 0.1846

Effect of learning condition on explicit solution: Besides subjective self-report, training condition also affected the objective measure of whether a participant was able to state a wholly- or partially-correct rule: *Feedback* participants were significantly more likely to do so than *No-Feedback* participants (Table 3).

Table 3. Statement of wholly- or partly-correct rule as a function of training group, Experiment 1.

	Correct rule	
Training group	FALSE	TRUE
No-Feedback	84	14
Feedback	66	27

χ^2 = 5.3116, df = 1, p = 0.02118

	At least partly-correct rule	
Training group	FALSE	TRUE
No-Feedback	78	20
Feedback	56	37

χ^2 = 7.6566, df = 1, p = 0.005656

Uni- vs. bimodality of generalization performance: The test-phase performance of participants who stated a (correct or incorrect) rule was bimodally distributed in both the *Feedback* and the *No-Feedback* conditions, with one mode near 1 (corresponding to those who stated the correct rule) and a much smaller mode near 0.5 (corresponding to those who stated an incorrect rule) (Figure 2, left side). This is consistent with what Kurtz et al. (2013) found for visual pattern-learning in situations that encourage rule use. *Feedback* participants who stated no rule showed unimodally-distributed performance with a peak near 0.5. Thus, in the *Feedback* condition, high generalization performance was almost exclusively confined to correct rule-staters.

In the *No-Feedback* condition, however, generalization performance was bimodal for non-rule-staters as well, again with modes near 1 and 0.5 (Figure 2, right side). This was unexpected, as bimodality is associated with rule learning. Closer inspection found that the two modes in fact corresponded to different features.

For 9 of the 19 near-perfect ($> 90\%$ correct) non-rule-stating *No-Feedback* participants, the pattern hinged on whether the consonants were fricatives vs. stops, a feature which participants found easy to recognize but hard to verbalize (e.g., "just went with words that ended with the same sound that the other words ended with"). High generalization performance in the *No-Feedback* condition could thus be achieved with or without explicit rule-stating.[1]

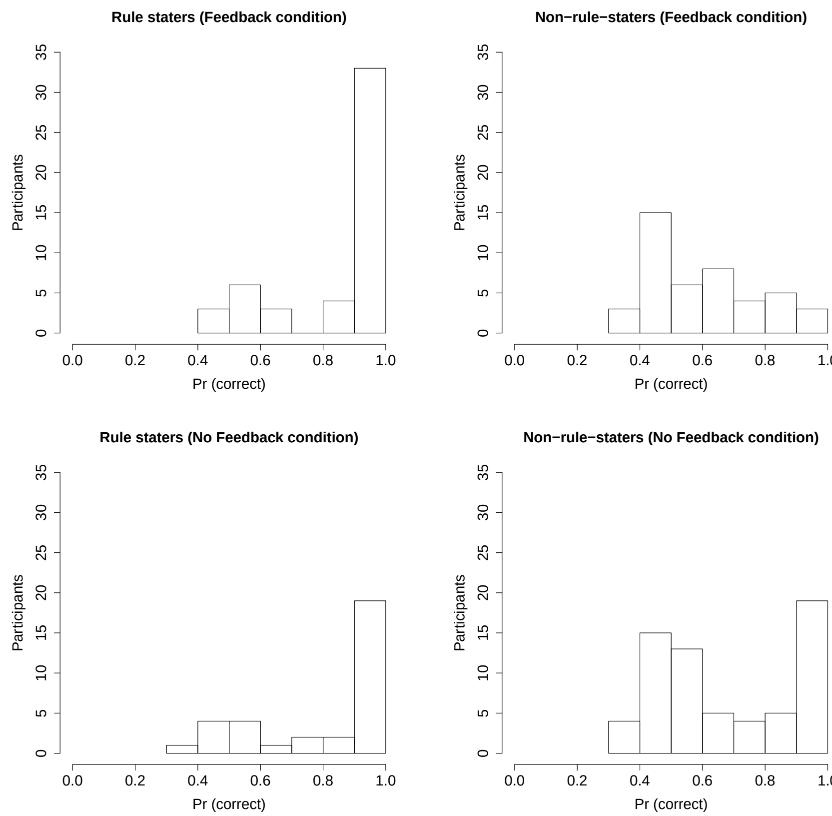


Figure 2. Uni- vs. bimodality of test-phase performance, Experiment 1.

Abruptness of solution: In the *Feedback* condition, the response to each training trial is correct or incorrect, making it possible to monitor progress during training. For each "solver" (*Feedback* participant who eventually reached the 16-consecutive-correct criterion), their last erroneous response before the 16-trial criterion run was located. The 16 trials preceding but not including the last error,

[1]The results are similar if the data is broken down by self-reported rule-seeking or self-reported rule-use rather than by rule-stating.

and the 16 trials following it, were extracted and divided into 4-trial blocks. Proportion correct was calculated for each of those blocks. If a block was incomplete because the last error occured before Trial 16, the proportion for that block was computed over its existing trials. If a whole block was missing for the same reason, no value was recorded for it. These individual means were then averaged together across participants (ignoring missing blocks) to yield the learning curves in Figure 3, which shows that improvement to criterion was more abrupt for those who stated the correct rule than for those who did not.

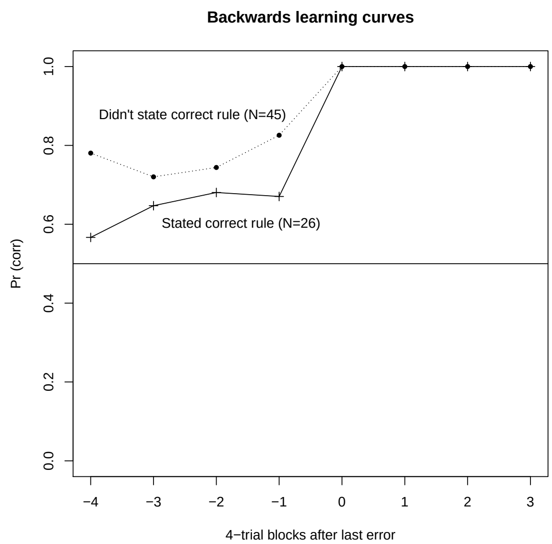


Figure 3. Backwards learning curves for the *Feedback* condition of Experiment 1. See text for details.

The difference in abruptness was tested by fitting a mixed logistic-regression model to the last 16 trials preceding each solver's last error (or fewer than 16, if the participant was a faster solver). The dependent variable was correct (1) vs. incorrect (0) on each trial. The independent variable was whether the participant was scored as correctly stating the rule (1) or not (0). There was a random intercept for each participant. The fitted model is shown in Table 4. The effect of *Stated correct* was significantly negative, i.e., just before they reached criterion, participants who stated the correct rule had worse performance than those who did not state the correct rule.

Table 4. Fixed effects part of logistic regression model for performance on last 16 pre-criterion trials by solvers in Experiment 1.

	Estimate	SE	z	$\Pr(>\|z\|)$	
Intercept	1.3195	0.1467	8.997	$< 2 \times 10^{-16}$	***
Stated correct	-0.5574	0.2568	-2.170	0.03	*

Response-time acceleration. Figure 4 shows that in the training phase of the Feedback condition, response times for correct responses accelerated after the last error for solvers who stated the correct rule, but not for solvers who didn't.

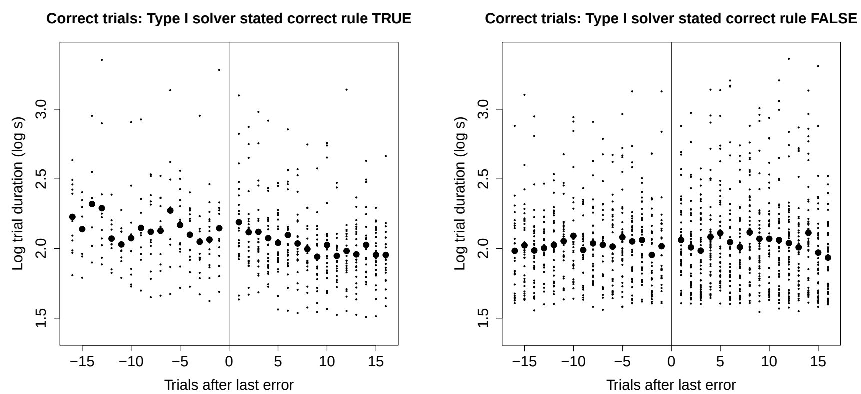


Figure 4. Response latencies for correct responses by *Feedback* solvers in the training phase of Experiment I, aligned to last error. Each small dot is one participant's log response time; large dots are means of log response times.

The acceleration after the last error was tested using a linear mixed-effects model in which log trial duration was the dependent variable. The independent variables were *After last error* (coded as -1 or $+1$ for trials before or after the last error), *Stated correct* (1 if the participant stated the correct rule, else 0), their interaction, and the log of the absolute trial number (to take into account the overall RT acceleration with practice). There was a random intercept for each participant. The software was the *lmer* function in R (Bates et al., 2015).

The significant interaction *After last error* $\times$ *Stated correct* (Table 5) confirms that those who stated the rule correctly experienced an acceleration in RTs after their last error. That could be because before the last error, they were seeking or testing a rule (slowly and effortfully), but after the last error they were applying a rule (quickly).

Table 5. Fixed effects part of general linear model for response times on correct trials by solvers in Experiment 1.

	Estimate	SE	df	t	$\Pr(>\|t\|)$	
Intercept	2.759	0.05166	180	53.418	< 0.001	***
Log trial number	-0.166	0.00903	4518	-18.407	< 0.001	***
After last error	-0.004	0.01033	4600	-0.369	0.7121	
Stated correct	-0.077	0.06949	70	-1.108	0.2717	
After last error $\times$ *Stated correct*	-0.046	0.01832	4617	-2.215	0.0268	*

1.3 Interim summary: Experiment 1

Pattern-learning in Experiment 1 shares some important properties with non-linguistic pattern-learning. Implicit and explicit processes are used by learners in both the *No-Feedback* and the *Feedback* condition. Changing the experimental task and instructions influences but does not determine the strategy; i.e., it is unsafe to assume that a particular experimental paradigm elicits only one kind of learning. Feedback was associated with more rule seeking, less report of intuition and memorization, and better performance. Solvers in the Feedback condition who stated the correct rule also displayed signs of rule- rather than cue-learning: an abrupt "aha" moment, followed by the speeding up of responses.

2 Experiment 2

Do implicit and explicit processes differ in their sensitivity to different patterns? Non-linguistic research has centered on "Type II" and "Type IV" patterns, in the terminology of Shepard et al. (1961) (Figure 5).

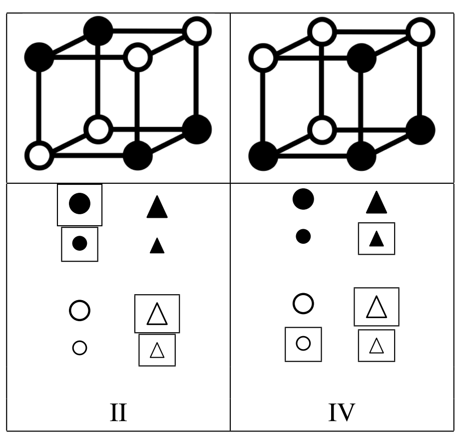


Figure 5. Top panels: Structure of patterns in a cube-like space defined by three binary dimensions. Bottom panels: Patterns instantiated in dimensions

black/white, circle/triangle, large/small. After Shepard et al. (1961).

The usual findings are that Type II (if-and-only-if) is easier than Type IV (family-resemblance) in terms of trials to criterion and total number of errors during training, and that switching from explicit to implicit learning reduces performance on Type II relative to Type IV (Love, 2002; Kurtz et al., 2013). Several proposals have been advanced in the psychology literature to account for the Type II $>$ Type IV advantage, based on the idea that (a) explicit rule learning is biased towards hypotheses that involve fewer features, and (b) only two features are relevant for Type II, while three are relevant for Type IV (for reviews see Pape et al. 2015; Moreton et al. in press). Experiment 2 therefore asks whether explicit processes also facilitate learning of Type II over Type IV in phonology.

2.1 Stimuli, methods, and participants

Of the six phonological properties in Experiment 1, we chose three that had elicited high test-phase performance in Experiment 1 (two/three syllables, labials/alveolars, fricatives/stops) as the three binary dimensions defining the stimulus space. Participants, recruited as in Experiment 1, were assigned to one of six groups defined by (Type I, Type II, Type IV) $\times$ (*No-Feedback*, *Feedback*). 167 people participated. Exclusions were: 6 reported non-English L1, 8 reported choosing test-phase responses that were maximally *un*like training, 14 reported note-taking, 2 fell below the 10-out-of-32 criterion. That left 137 valid participants, between 16 and 28 in each of the 6 cells. Questionnaire responses were scored by the two authors. Inter-rater reliability was at least 89% for each question. Disagreements were resolved by the first author, comparing raters' notes. Since Type II and Type IV rules were hard to completely verbalize, a stated rule was scored as *correct* (gave correct answers for every stimulus), *partly correct* (gave correct answers for more than half of stimuli), or *incorrect* (other).

2.2 Results

For space reasons, only the Type II and Type IV conditions are analyzed here. (The Type I condition's results followed the pattern of Experiment 1, though with less statistical power owing to the smaller sample.)

Trials to criterion and total training errors. Unlike in the non-linguistic literature, in the *Feedback* condition, solvers reached criterion significantly *earlier* for Type IV than for Type II (Wilcoxon Mann-Whitney rank-sum test, $Z = 2.0392, p = 0.04143$), and made significantly *fewer* errors during training ($Z = 2.0671, p = 0.03872$). There were no significant effects of, nor interactions with, rule-seeking or rule-stating.

Generalization: Non-linguistic studies have found that when explicit rule-seeking is encouraged, the Type II advantage over Type IV increases (Love, 2002; Kurtz et al., 2013). In the test phase of Experiment II, however, the opposite happened: Type IV was better than Type II, and the Type IV advantage was greater for those who reported seeking a rule in the training phase. The statistical analysis (using the *glmer* function in the *lme4* package of R, Bates et al. 2015) is shown in Table 6. The dependent variable is correct/incorrect response (1/0). The fixed effects are *Sought* (=1 if reported rule-seeking, else 0), *IV* (= 1 for participants in Type IV, 0 for Type II), *No-Feedback* (=1 for *No-Feedback* participants, 0 for *Feedback* participants). A random intercept was included for each participant. The significant and positive *Sought*×*IV* interaction, with the non-significant main effects of these two variables, means that in the *Feedback* Type IV condition, but not in the *Feedback* Type II condition, rule-seeking significantly increased the probability of a correct response.

Table 6. Fixed effects part of logistic-regression model for test phase of Experiment 2.

	Estimate	SE	z	$\Pr(>\lvert z\rvert)$	
(Intercept)	0.8810	0.4100	2.149	0.0317	*
Sought	−0.1454	0.4584	−0.317	0.7511	
IV	−0.6338	0.5329	−1.189	0.2343	
No-Feedback	−0.0140	0.5554	−0.025	0.9799	
Sought:*IV*	1.4725	0.6085	2.420	0.0155	*
Sought:*No-Feedback*	0.1658	0.6415	0.258	0.7960	
IV:*No-Feedback*	0.5652	0.7186	0.786	0.4316	
Sought:*IV*:*No-Feedback*	−1.1190	0.8868	−1.262	0.2070	

2.3 Discussion: Experiment 2

Contrary both to the results of most non-linguistic experiments, and to the predictions of models that account for them, performance was *better* on Type IV than Type II, and explicit rule-seeking was associated with an *increase* in the Type IV advantage. A possible explanation is as follows: The two or three relevant features in Exp. 2 had to be discovered amongst seven or six irrelevant ones ("attribute identification", Haygood & Bourne 1965). Suppose rule-seekers do that by sequentially testing individual features for cue validity. Each of the three Type IV features, by itself, allows 75% correct responding. But for Type II, a single relevant feature is useless — wrong 50% of the time — and so cannot be distinguished from an irrelevant feature. That makes it harder to find the solution incrementally.[2,3] This conjecture is supported by the relation between the Type

[2]It is not the case that Type IV participants were using only single-feature rules. If many participants were using a partly-correct single-feature rule, then the test-phase responses should have had a mode at 75% correct. In fact, the distribution of responses in both the *Feedback* and *No-Feedback* conditions had a valley near 75%, with modes above and below.

[3]Type IV is linearly separable in terms of the features, while Type II is not, but linear separability does not appear to be at the root of the Type IV advantage: In a study using isomorphic phonotactic and visual patterns, Type II also proved harder than a three-feature non-linearly-separable pattern as well (Moreton et al., in press).

conditions and rule-seeking and -stating, shown in Table 7.

Table 7. Report of (at least partly) correct rule in Experiment 2 as a function of Type and self-reported rule-seeking.

	Type II		Type IV	
Stated (some) correct	Nonseekers	Seekers	Nonseekers	Seekers
FALSE	10	30	16	13
TRUE	1	4	0	15

As shown in Table 7, the only group in which most participants reported a rule that was at least partly correct was rule-seekers in the Type IV condition. In the other three groups, almost no one reported such a rule. (The seeker/nonseeker difference within the Type IV group was significant by Fisher's exact test, $p = 0.001897$; likewise, the II/IV difference within the seekers was also significant by the same test, $p = 0.0006718$).

3 General discussion

Experiment 1 found evidence that participants in "artificial-language" experiments can use (at least) two qualitatively distinct learning processes, explicit rule-seeking and implicit intuitive learning. The experimenter's choice of task and instructions can favor the use of either process, but does not guarantee that one process will be used exclusively. Explicit and implicit learners differ in multiple ways, as measured by self-report of strategy, explicit rule statement, abruptness of learning curves, response-time acceleration, and uni- vs. bimodality of generalization performance. Experiment 2 found that differences in learning strategy can lead to difference in relative pattern difficulty.

Thus, that there can be two sub-populations of learners doing different things in the same experiment, whose behaviors may cancel each other out when aggregated. Disaggregation may therefore reveal or strengthen effects which have hitherto been weak or hard to replicate (e.g., the elusive evidence for "phonetically substantive" inductive bias, Moreton & Pater 2012a,b).

Experiment 2 also found that rule-seeking is associated with an increase in the $IV > II$ advantage, apparently because it is easier to find the relevant features one by one in Type IV. This is a new finding, unexpected under psychological models of rule-based learning which favor patterns that depend on fewer features (see references above). Feature-minimizing inductive biases have also been independently proposed in phonology (e.g., Chomsky & Halle 1968, 168, 221, 331, 334; Bach & Harms 1972; Pycha et al. 2003; Gordon 2004; Hayes et al. 2009); they, too, are challenged by the present results. Constraint-based models do exist which predict $IV > II$ (Gluck & Bower, 1988; Pater & Moreton, 2012; Moreton et al., in press), but they do not predict an amplifying effect of deliberate rule-seeking. New modelling ideas are clearly needed.

Explicit learning is often regarded as a contaminant ("using a strategy", "solving crossword puzzles", etc.), and these results may be of interest to experimenters desiring to minimize it. However, what adult "artificial-language" studies resemble most is the early stages of *second*-language acquisition, which can make use of both implicit and explicit processes (reviewed in Lichtman 2012); hence, explicit phonological learning is also worth studying in its own right.

References

ASHBY, F. GREGORY; LEOLA A. ALFONSO-REESE; AND U. TURKEN; and ELLIOTT M. WALDRON. 1998. A neuropsychological theory of multiple systems in category learning. *Psychological Review* 105.442–481.

BACH, EMMON, and ROBERT T. HARMS. 1972. How do languages get crazy rules? *Linguistic change and generative theory*, ed. by R. P. Stockwell and R. K. S. Macaulay, chap. 1, 1–21. Bloomington: Indiana University Press.

BATES, DOUGLAS; MARTIN MÄCHLER; BEN BOLKER; and STEVE WALKER. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48.

BRUNER, JEROME S.; JACQUELINE J. GOODNOW; and GEORGE A. AUSTIN. 1956. *A study of thinking*. New York: John Wiley and Sons.

CHOMSKY, NOAM, and MORRIS A. HALLE. 1968. *The sound pattern of English*. Cambridge, Massachusetts: MIT Press.

CIBOROWSKI, TOM, and MICHAEL COLE. 1972. A cross-cultural study of conjunctive and disjunctive concept learning. *Child Development* 43.774–789.

CIBOROWSKI, TOM, and MICHAEL COLE. 1973. A developmental and cross-cultural study of the influences of rule structure and problem composition on the learning of conceptual classifications. *Journal of Experimental Child Psychology* 15.193–215.

EVANS, J. S. B. T. 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology* 49.255–278.

GLUCK, MARK A., and GORDON H. BOWER. 1988. Evaluating an adaptive network model of human learning. *Journal of Memory and Language* 27.166–195.

GORDON, MATTHEW. 2004. Syllable weight. *Phonetically-based phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 277–312. Cambridge, England: Cambridge University Press.

HAIDER, HILDE, and MICHAEL ROSE. 2007. How to investigate insight: a proposal. *Methods* 42.49–57.

HAYES, BRUCE; KIE ZURAW; PÉTER SIPTÁR; and ZSUZSA LONDE. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85.822–863.

HAYGOOD, R. C., and L. E. BOURNE. 1965. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review* 72.175–195.

KURTZ, KENNETH J.; KIMERY R. LEVERING; ROGER D. STANTON; JOSHUA ROMERO; and STEVEN 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* 39.552–572.

LEWANDOWSKY, STEPHAN. 2011. Working memory capacity and categorization: individual differences and modelling. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37.720–738.

LICHTMAN, KAREN MELISSA. 2012. *Child-adult differences in implicit and explicit second language learning*. University of Illinois at Urbana-Champaign dissertation.

LOVE, BRADLEY C. 2002. Comparing supervised and unsupervised category learning. *Psychonomic Bulletin and Review* 9.829–835.

MADDOX, W. TODD, and F. GREGORY ASHBY. 2004. Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes* 66.309–332.

MORETON, ELLIOTT. 2012. Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language* 67.165–183.

MORETON, ELLIOTT, and JOE PATER. 2012a. Structure and substance in artificial-phonology learning: Part I, structure. *Language and Linguistics Compass* 6.686–701.

MORETON, ELLIOTT, and JOE PATER. 2012b. Structure and substance in artificial-phonology learning: Part II, substance. *Language and Linguistics Compass* 6.702–718.

MORETON, ELLIOTT; JOE PATER; and KATYA PERTSOVA. in press. Phonological concept learning. To appear in *Cognitive Science*.

NEWELL, B. R.; J. C. DUNN; and M. KALISH. 2011. Systems of category learning: fact or fantasy? *The psychology of learning and motivation*, ed. by B. Ross, vol. 54, chap. 6, 167–215. Academic Press.

NOSOFSKY, ROBERT M., and THOMAS J. PALMERI. 1996. Learning to classify integral-dimension stimuli. *Psychonomic Bulletin and Review* 3.222–226.

OSMAN, M. 2004. An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin and Review* 11.988–1010.

PAPE, ANDREAS D.; KENNETH J. KURTZ; and HIROKI SAYAMA. 2015. Complexity measures and concept learning. *Journal of Mathematical Psychology* 64–65.66–75.

PATER, JOE, and ELLIOTT MORETON. 2012. Structurally biased phonology: complexity in learning and typology. *Journal of the English and Foreign Languages University, Hyderabad* 3.1–44.

PERTSOVA, KATYA. 2012. Logical complexity in morphological learning. To appear in *Proceedings of the Berkeley Linguistics Society*.

PYCHA, ANNE; PAWEL NOWAK; EURIE SHIN; and RYAN SHOSTED. 2003. Phonological rule-learing and its implications for a theory of vowel harmony. *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)*, ed. by M. Tsujimura and G. Garding, 101–114.

SHEPARD, R. N.; C. L. HOVLAND; and H. M. JENKINS. 1961. Learning and memorization of classifications. *Psychological Monographs* 75.

SMITH, J. DAVID; MARK E. BERG; ROBERT G. COOK; MATTHEW S. MURPHY; MATTHEW J. CROSSLEY; JOE BOOMER; BRIAN SPIERING; MICHAEL J. BERAN; BARBARA A. CHURCH; F. GREGORY ASHBY; and RANDOLPH C. GRACE. 2012. Implicit and explicit categorization: a tale of four species. *Neuroscience and Biobehavioral Reviews* 36.2355–2369.

SMITH, J. DAVID; JOHN PAUL MINDA; and DAVID A. WASHBURN. 2004. Category learning in rhesus monkeys: a study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General* 133.398–404.

SPROUSE, JON. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgements in linguistic theory. *Behavior Research Methods* 43.155–167.