The experimental study of artificial language learning has become a widely used means of investigating the predictions of theories of phonology and of learning (see Moreton and Pater (2012a; 2012b) for a review). Although much is now known about the generalizations that learners make from various kinds of data, relatively little is known about how those generalizations are cognitively encoded. Specifically, are generalizations learned in the lab comparable to generalizations learned as part of naturalistic language acquisition?

Models of phonological knowledge fall into two broad classes: lexical (analogical) vs. abstract (grammatical) (see e.g. Albright and Hayes 2003; Ernestus and Baayen 2003; Daland et al. 2011). This paper provides evidence that generalizations acquired in the lab can be encoded at an abstract level. Brain responses to violations of lab-learned phonotactics were examined. Novel words that violated a learned phonotactic constraint elicited a larger Late Positive Component (LPC; also known as P600) than novel words that satisfied it. The LPC is also associated with syntactic violations and with violations of musical expectations (overview in Gouvea et al. 2010) - knowledge which is considered to be abstract in nature. Domahs et al. (2009) also found an LPC in response to violations of participants' native language phonotactics. Based on this previous literature, we argue that the presence of an LPC in response to a lab-learned phonotactic pattern constitutes evidence that that pattern (a) is abstractly represented, similarly to a syntactic generalization, and (b) is represented in a cognitively similar way to native language phonotactics.

**Methods:** We taught 24 adult, native English speakers 16 novel words with meanings. They were asked to match an auditorily presented word to 1 of 4 pictures, after which they were shown the correct picture. The words each participant learned were all consistent with a phonotactic pattern, which was a consonant voicing agreement pattern for Group 1, and a consonant voicing disagreement pattern for Group 2. Each group contained 12 participants. All stimuli were CVCV words. The patterns, and the stimulus space, were as studied by Moreton (2008; 2012). The stops were drawn from the set [d, g, t, k], and the vowels from the set [i, æ, u, ɔ]. We constructed 48 words, with the consonants in half of them agreeing in voice (e.g. [dugi], [tikɔ]), and in the other half disagreeing (e.g. [kædu], [tigæ]). Participants were trained on items following the pattern they were learning (either agreement or disagreement). During testing, participants were asked to rate on a 4-point scale how likely it is that each word is part of the language they were learning. These words included eight trained words, eight novel words that fit the pattern (Untrained Fit), and eight words that violated it (Untrained Unfit). Testing and training blocks alternated, with a total of five each.

**Results:** During testing, participants rated Untrained Fit words as more likely to be in the language (M = 2.71, SD = 0.29) than Untrained Unfit words (M = 2.21, SD = 0.26) (t(23) = 7.98, p < 0.001). Untrained Unfit words elicited a larger LPC than Untrained Fit
words. Over central-posterior and posterior regions, novel words that did not fit the pattern of the trained language elicited a larger negativity 600-1000 ms after onset (F(1,23) = 5.55, p = 0.027). The figure shows grand average ERPs to novel words which fit the trained pattern (Untrained Fit, black) and to novel words which violate the trained pattern (Untrained Unfit, grey), as well as the scalp distribution of the positivity. The four electrodes shown are from posterior and central regions, and are indicated on the scalp map to the right with stars. The scalp map shows the average voltage over the time window 600-1000 ms post-stimulus-onset. Redder colors are more positive, and bluer colors are more negative. Units are µV.

In addition to the LPC, an N400 effect was observed between trained items and untrained items. The N400 is often interpreted as an index of the effort involved in lexical access (e.g. Coch et al. 2015). In our experiment, Untrained-Fit items elicited a greater negativity than Trained items over central and posterior regions between 400 and 700 ms post stimulus onset (F(1,23) = 6.67, p = 0.017). No difference in N400 size was found between Untrained-Fit and Untrained-Unfit items. Such a difference would indicate that participants were using a lexical search strategy, such as analogy, to perform the rating task.

**Discussion:** The participants in our experiment learned a dependency between the voicing of the two stop consonants of CVCV words. They were exposed to a set of words obeying the restriction in the context of learning the meanings of the words, and then in testing they rated novel words that fit the restriction as more likely to belong to the language than novel words that violated it. From the study of EEG data collected during the experiment, we conclude that participants learn an abstractly represented phonotactic generalization. The ERP response to the Untrained-Unfit items included a Late Positive Component (LPC), similar to that found in response to syntactic and musical harmonic structure violations, and similar to the late positivity found previously for violations of natively learned phonotactic generalizations. This finding lends credibility to the use of artificial language learning experiments in the investigation of human language acquisition.