

# Appendix A

## Study 1 Power analysis

### A.1. Pilot study

Prior to Experiment 1, a pilot study was run that included three of the four conditions included in Experiment 1A: Native Shared (2 English talkers), Weak Different (1 English talker and 1 French talker with the [u] vowel), and Strong Different (1 English talker and 1 French talker with the [y] vowel). Other than the exclusion of the Non-Native Shared condition, the stimuli, design, and procedure of the pilot were identical to those of Experiment 1A. The power analysis for Experiment 1 was based on the results from this pilot.

### A.2. Participants

To reach the target of 48 participants (16 per condition) who passed the experimental criteria, a total of 85 native speakers of English were recruited on AMT (passing rate of 56.5%).

### A.3. Data analysis

Data analysis was identical to that in Experiment 1A with one exception: The accent term, comparing the two Shared conditions, was not included, as there was only one shared condition. The pilot data was analyzed using a logistic mixed-effects regression, with participant responses as the dependent measure. Fixed effects included *legality*, and two contrast-coded terms: *language difference* (i.e., Shared versus Different conditions), and *strength* (i.e., Weak versus Strong Different conditions). Interaction terms were included between legality and both contrast-coded terms. Random effects included random intercepts and random slopes by legality for both participants and items. In addition, follow-up analyses were run on individual conditions, which included a fixed effect of legality, and random intercepts, as well as slopes by legality for items (the models did not converge with random slopes by participant).

### A.4. Pilot results

The analysis revealed a main effect of legality ( $\beta = 0.49$ ,  $SE \beta = 0.12$ ,  $\chi^2(1) = 15.23$ ,  $p < 0.001$ ), suggesting that participants were, overall, able to learn the constraint. In addition, there was a marginal interaction between the shared term and legality ( $\beta = -0.63$ ,  $SE \beta =$

0.34,  $\chi^2(1) = 3.31, p = 0.07$ ), providing weak evidence that participants adapted to a greater degree in the Different conditions. This was consistent with a follow-up analysis showing that there was a significant difference between participant responses on legal and illegal syllables in both of the Different conditions (Strong:  $\beta = 0.73, SE \beta = 0.19, \chi^2(1) = 13.1, p < 0.001$ ; Weak:  $\beta = 0.46, SE \beta = 0.19, \chi^2(1) = 6.14, p < 0.05$ ), but no such difference in the Shared Native condition ( $\beta = 0.19, SE \beta = 0.20, \chi^2(1) = 0.92, p = 0.34$ ). This suggested that listeners only adapted to talker-specific phonotactic constraints if speakers differed in their language background, as was found in Experiments 1 and 2. There was no significant interaction between strength and legality ( $\beta = 0.32, SE \beta = 0.30, \chi^2(1) = 1.16, p = 0.28$ ), suggesting both conditions provided sufficient cues for listeners to identify a difference in language background, similar to Experiment 1.

#### A.5. Power analysis

The number of participants was set to yield sufficient statistical power ( $\beta > .8$ ). Power was estimated by Monte Carlo simulations based on the results above. Using the estimates for each fixed and random effect in the logistic mixed effects model fit to these pilot data, we generated 1000 simulated data sets. For each simulated data set, we randomly and independently sampled each fixed effect value from a normal distribution (with the mean set to the respective coefficient estimate and standard deviation set to the corresponding standard error estimate) and independently sampled each random effect based on the estimated random effect distributions (correlations between coefficients were not incorporated into our sampling procedure). These were then used to generate a set of recognition memory test response. We then fit the same regression model to these simulated responses. (If the model failed to converge, we generated a new simulated data set.) Statistical power  $\beta$  was estimated by the proportion of the 1000 models in which the crucial interaction term—between the fixed effects for legality and shared/different language background—was found to be significant. We increased the number of participants iteratively, generating novel simulated data sets and running new models with each iteration, until we reached the threshold of  $\beta > .8$ . This threshold was reached with 64 participants per condition (estimated  $\beta = .804$ ). The scripts and data for the power analysis can be found on the Open Science Framework platform (<https://osf.io/et9jd/>).