

Appendix B

The impact of changes to participant inclusion criteria on the results of study 1

Recall that experimental criteria in Experiments 1A and 1B were as follows: In the generalization phase, participants had to correctly accept at least 90% of previously heard items (i.e., a hit rate of over 90%) and correctly reject at least 10% of novel items. These criteria ensured that participants were able to recall items they had previously heard multiple times, and that they were able to differentiate between previously heard items and novel items. In Experiment 2, these criteria were loosened due to the phonological confusability of the stimulus set: The hit rate criterion was lowered from 90% to 85%. A third criterion was also added: Listeners' hit rate could not exceed their false alarm rate, to ensure that listeners could differentiate novel and familiar items. For example, a participant with a hit rate of 85% must correctly reject at least 15% of novel items (i.e., a false alarm rate no higher than 85%).

Table B.1. Passing rates for each condition.

Experiment	Condition	Total participants	Passing Participants	Passing Rate
1A	Native Shared	124	64	51.6%
1A	Non-Native Shared	101	66	65.3%
1A	Weak Different	118	66	55.9%
1A	Strong Different	112	64	57.1%
1B	Non-Native Shared	158	66	41.8%
1B	Weak Different	119	64	53.8%
1B	Strong Different	141	64	45.4%
2	Non-Native Shared	160	66	41.3%
2	Non-Native Different	157	67	42.7%
2	Mixed Different	124	69	55.6%

N.B. The same number of participants (64) were analyzed in each condition. The number of passing participants sometimes exceeded this due to technical limitations in our experimental pipeline.

As shown in Table B.1, passing rates ranged from 41.3% to 65.3% between conditions. Overall, 655 participants passed the criteria out of 1314 participants (49.8%). This was in line

with previous results using this paradigm (Denby, Schechter, Arn, Dimov, & Goldrick, 2018). In a post-hoc analysis, we investigate the relationship between the hit rate in the generalization phase and the legality effect. As Figure B.1 shows, participants with a hit rate lower than roughly 75% show little to no legality effect. This is unsurprising: If, for example, a participant correctly accepts only half of familiar items, they are simply at chance, and therefore will not show any differences between legal and illegal generalization items, as they are likely guessing. Participants whose hit rate is much lower than 50% may have misinterpreted the experimental instructions, and simply answered “no” to any items they had not encountered prior to the experiment (rather than within the experiment).

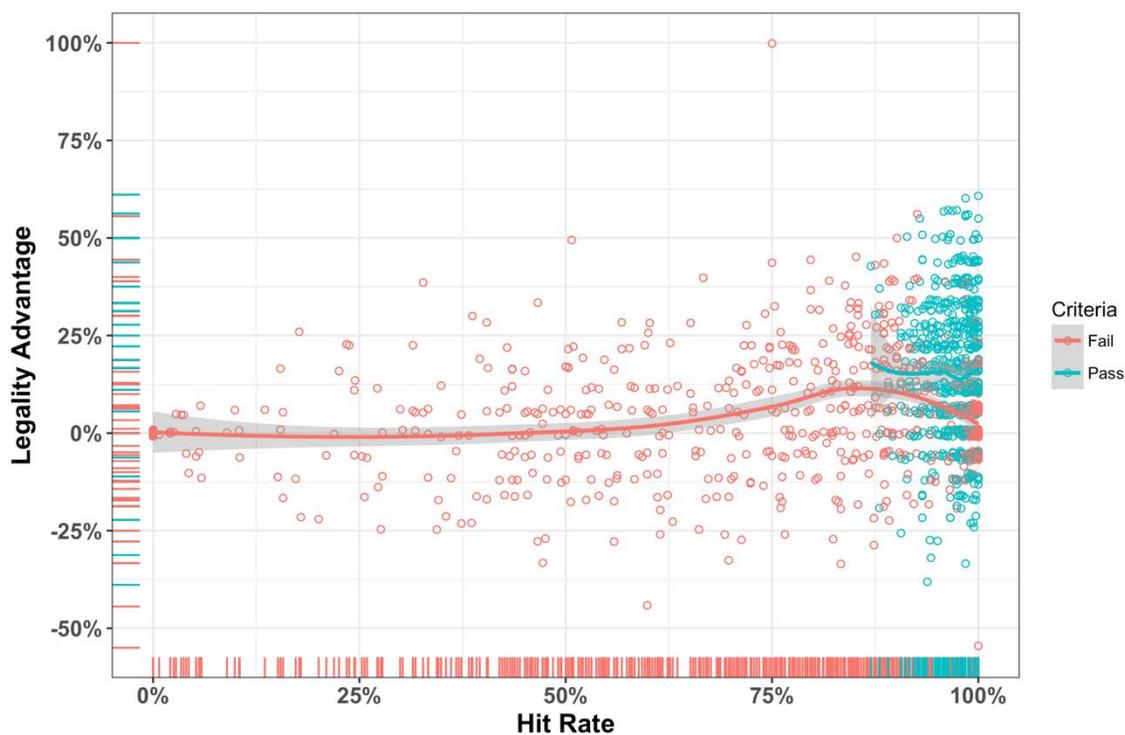


Figure B.1. Scatterplot of hit rate (% yes on familiar items) in generalization phase by legality advantage (false alarm rate for legal items minus false alarm rate on illegal items) for all experiments in Study 1. Each dot represents a single participant; colors represent whether participants passed or failed criteria. Lines represent Loess regression; shading represents 95% confidence interval.

Note that the relationship between hit rate and legality advantage is non-linear—for participants who fail the criteria, the legality advantage peaks around a hit rate of 85%; as the hit rate increases to 100%, the legality advantage falls back down to almost 0%. This is also an

expected result, as such participants are failing the criteria based on a high false alarm rate: They are responding “yes” to almost every item, regardless of whether it is familiar or novel. Detecting differences in response patterns between legal and illegal novel items is essentially impossible with such a high overall false alarm rate.

These results suggest that the criteria were necessary to filter out participants who were biased towards always responding “yes” or always responding “no,” as well as those who answered randomly (i.e., responded “yes” ~50% of the time). It also appears as though the criteria may have been slightly too restrictive, as participants whose hit rate was above roughly 75% appeared to be tracking the constraint, as shown by their increased legality advantage. In a second post-hoc analysis, we re-plotted the data while loosening the criteria to include participants with a hit rate as low as 75%. As in Experiment 2, we included a criterion that participants’ false alarm rate must be lower than their hit rate, to ensure they are able to differentiate familiar and novel items.

Loosening the criteria resulted in an additional 152 participants passing, increasing the overall passing rate from 49.8% to 61.4%. As can be seen in Figures B.2 – B.4, the results of the experiment do not qualitatively change with the addition of these participants. This suggests the criteria as originally set were somewhat overly restrictive, with an additional ~10% of participants unnecessarily excluded. Based on these results, we recommend that future experiments with similar designs should loosen the criteria to 75%.

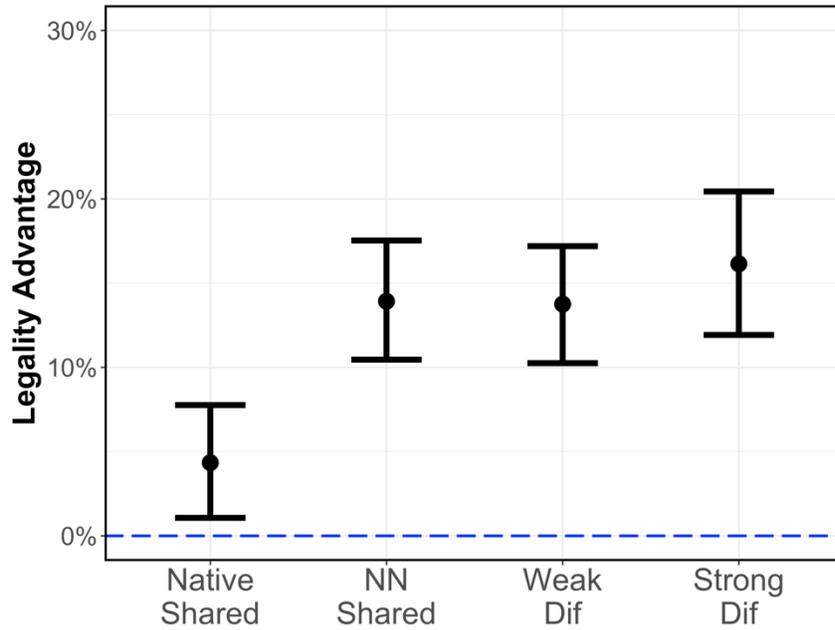


Figure B.2. Legality advantage for Experiment 1A, with hit rate criterion lowered to 75%. Error bars reflect bootstrapped 95% confidence interval.

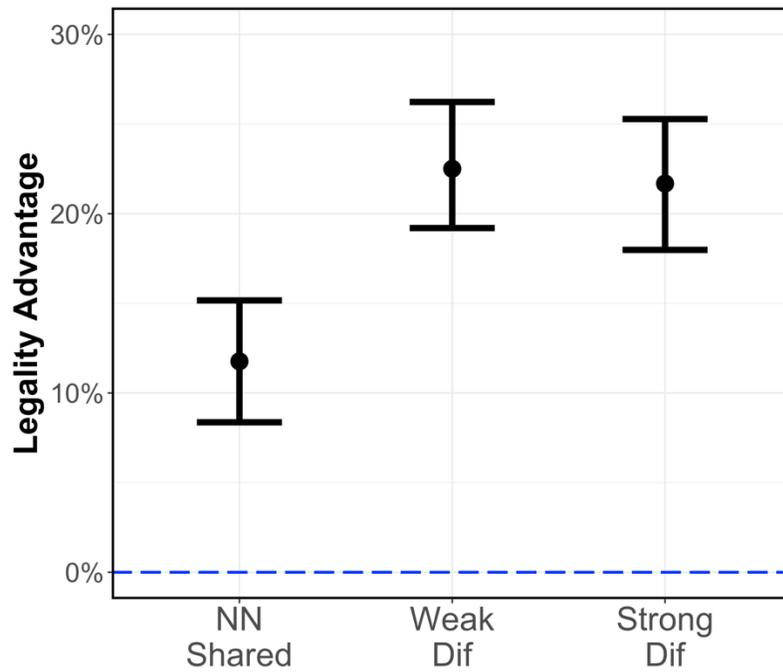


Figure B.3. Legality advantage for Experiment 1B, with hit rate criterion lowered to 75%. Error bars reflect bootstrapped 95% confidence interval.

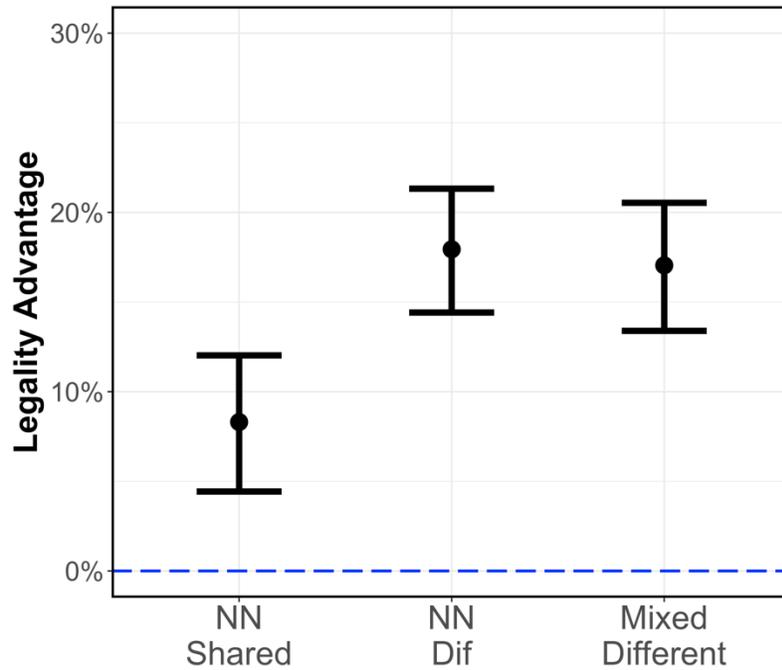


Figure B.4. Legality advantage for Experiment 2, with hit rate criterion lowered to 75%. Error bars reflect bootstrapped 95% confidence interval.

References

Denby, T., Schechter, J., Arn, S., Dimov, S., & Goldrick, M. (2018). Contextual variability and exemplar strength in phonotactic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(2), 280–294. <http://doi.org/10.1037/xlm0000465>