Vowel harmony and disharmony are not equivalent in learning

Alexander Martin

Centre for Language Evolution

University of Edinburgh

James White

University College London

1 Introduction

There has been significant debate about which analytic biases influence phonological learning and might therefore shape the development of phonological systems over time. For instance, substance-based theories of phonology posit that phonetically-motivated biases in the minds of speaker-listeners influence how they acquire phonological grammars (e.g., Archangeli & Pulleyblank, 1994; Donegan & Stampe, 1979; Hayes & Steriade, 2004), but others argue that phonetic substance should be limited to the diachronic domain (e.g., Blevins, 2004; Ohala, 1993).[1] Vowel harmony, along with its comparison to vowel dishar-

mony, has taken center stage in the literature on learning biases in phonology. General vowel harmony and disharmony rules have comparable formal complexity (see Moreton & Pater, 2012a), but differ dramatically in terms of their typological frequency and phonetic motivation. Vowel harmony occurs frequently (for an overview, see Rose & Walker, 2011) and has clear phonetic motivation (vowel-to-vowel coarticulation; Ohala, 1994);[2] vowel disharmony is exceedingly rare and has no clear phonetic motivation. This combination makes vowel harmony versus disharmony an excellent test case for exploring the role of biases in phonology.

Earlier experimental studies comparing the learnability of vowel harmony and vowel disharmony found no difference between the two, though they found that both patterns were easier to learn than more complex vowel co-occurrence patterns involving an arbitrary combination of features (Pycha, Nowak, & Shin, 2003; Skoruppa & Peperkamp, 2011).[3] Recent work has provided some evidence of better learning of vowel harmony than vowel disharmony when using non-native stimuli (Martin & Peperkamp, forthcoming), suggesting that substantive biases might emerge more strongly when listeners are required to tap into phonetic knowledge. Nevertheless, the earlier findings have formed part of the core support for the view that substantive learning biases are weak, at best, and possibly non-existent, whereas learning biases motivated by structural complexity are strong and robust (argued perhaps most explicitly in Moreton & Pater, 2012a, 2012b). Moreover, the assumption that there is no difference in learnability between vowel harmony and vowel disharmony leads to the conclusion that any typological difference between the two patterns must be due to factors external to learning (e.g., channel bias; Moreton, 2008).

In this paper, we argue that in fact all previous learning experiments comparing har-

mony and disharmony have missed a crucial distinction due to the nature of their test phases. Specifically, these studies assume that learners have acquired general rules (or constraints) of the type in (1); however, we show that when a more nuanced test is used, it becomes clear that learners do not readily infer the general disharmony pattern in (1b).

(1) General harmony and disharmony rules (constraints)

a. Harmony: $V \rightarrow [\alpha F]/[\alpha F]$ __ (or $*[\alpha F][-\alpha F]$)

b. Disharmony: $V \rightarrow [-\alpha F]/[\alpha F]$ __ (or $*[\alpha F][\alpha F]$)

If participants exposed to disharmony are truly acquiring a general vowel disharmony rule, as in (1b), then they should prefer forms with fully alternating feature values, such as $[+F][-F][+F][-F]$, compared to forms like $[+F][-F][+F][+F]$ with the alternating pattern disrupted. Participants acquiring a general vowel harmony rule (1a) should prefer forms with fully consistent feature values, $[+F][+F][+F][+F]$ and $[-F][-F][-F][-F]$, compared to ones like $[+F][+F][+F][-F]$. Previous studies, however, only tested participants on two-syllable stems (Skoruppa & Peperkamp, 2011) or on a stem plus a single suffix (Pycha et al., 2003; Martin & Peperkamp, forthcoming), making it impossible to know whether participants learned a general vowel (dis)harmony rule or some other (less general) pattern. With a single suffix, for instance, participants could succeed on a putatively disharmonic pattern by using a number of alternative strategies, such as learning agreement with a nonlocal vowel (e.g., see White et al., 2018) or learning a correspondence between one allomorph and certain stem types (e.g., "put *-fu* after stems ending in *e*").

In the current study, we exposed participants to stem + suffix forms that were consistent with a vowel harmony or vowel disharmony pattern (depending on condition). Participants were trained on forms with only a single suffix, but in the test phase we crucially required

them to extrapolate (without additional training) to forms with two suffixes, allowing us to gauge whether they had actually learned a general vowel harmony/disharmony pattern like in (1). To preview the results, we found that participants exposed to vowel harmony overwhelmingly inferred a pattern consistent with general vowel harmony. Participants exposed to disharmony, on the other hand, did not infer a general disharmony pattern. Our results strongly suggest that vowel harmony has a preferential status in learning compared to vowel disharmony, despite the formal similarity between the two patterns.

2 Experiment

We conducted an artificial language learning experiment using the "poverty-of-the-stimulus" paradigm (Wilson, 2006). In this paradigm, participants are trained on ambiguous input and then tested on disambiguating cases to see which pattern they have inferred. This paradigm has been used extensively to test for learning biases, in phonology (i.a., Finley & Badecker, 2008; Kimper, 2016; White, 2014; White et al., 2018; Wilson, 2006), syntax (i.a., Culbertson & Adger, 2014; Martin, Abels, Adger, & Culbertson, 2019; Martin, Ratitamkul, Abels, Adger, & Culbertson, 2019), and more recently in semantics (Maldonado & Culbertson, 2019). In our experiment, participants were trained on either a vowel harmony pattern (backness harmony) or a vowel disharmony pattern. Crucially, participants were only exposed to cases of stems with a single suffix during training. In the harmony condition, participants encountered harmonic stems with a harmonic suffix (e.g., *peti-be* or *peti-fi*). In the disharmony condition, participants were trained on disharmonic stems and a suffix disharmonic with the final stem vowel (e.g., *petu-be* or *petu-fi*). In the test phase,

participants were required to apply both suffixes at once, which they had never encountered before. Our methodology and analysis plan were pre-registered on the OSF platform (https://osf.io/emnqg) before data were collected.

2.1 Method

2.1.1 Participants

We recruited a total of 272 participants through Amazon's Mechanical Turk platform, requiring them to be "master workers" who were based in the US, were above the age of 18, and spoke English at home growing up. Participants were randomly assigned to one of the two conditions (harmony or disharmony). Participants who started the experiment more than once (e.g., by refreshing the page) were excluded (N=68).

To ensure data quality, we included two important attention checks. First, near the end of the instructions, participants were told that the following page would ask for their Mechanical Turk ID, but that they should write "I understand" in the box instead. Second, we included four attention-check trials during the experiment itself (described in the Procedure section below). Participants who failed to enter "I understand" (N=59) or who failed more than one attention trial (N=21) were excluded and replaced. Participants were also excluded if they had participated in a pilot version of the experiment (N=4).

In total, we analyzed data from 120 participants (60 per condition), the number we pre-registered. All participants (excluded or not) received compensation of 6 USD.

2.1.2 Materials

In the training phase, stems for the harmony condition consisted of 32 CVCV nonce words, where the two vowels were both front or both back (e.g., *peti*). Front vowels were {*i, e*} and back vowels were {*u, o*}; each possible vowel combination (e.g., CiCi, CiCe, etc.) occurred an equal number of times. Consonants were drawn from the set {*p, t, k, b, d, g, m, n, s, l*}, balanced across items and across word positions. Words that sounded like real English words were avoided. Each stem was paired with an image showing a singular object or animal. Stems for the disharmony condition were derived from the 32 harmonic stems by changing the backness of the second vowel (e.g., *peti* → *petu*); thus, the stems of the disharmony condition were identical to those in the harmony condition except that the two vowels disagreed in backness. For each training stem, suffixed forms were created by adding a -CV suffix. One suffix (-*be/-bo*) contained a mid vowel and one (-*fi/-fu*) contained a high vowel. Each suffix was assigned one of two meanings, *plural* or *diminutive*, counterbalanced across participants. Each stem occurred twice in the training phase, once with the -*be/-bo* suffix and once with the -*fi/-fu* suffix.

The testing phase consisted of 16 new CVCV stems, which were created following the same procedure described for the training phase above, yielding 32 suffixed forms. In addition, we created forms with both suffixes together for each stem (e.g., *peti-be-fi*).

All stimuli other than the attention-check items were recorded by a phonetically-trained male native speaker of American English (the second author). The attention-check items were recorded by a female native speaker of American English not involved in the study. Stimuli can be downloaded from the first author's website.

2.1.3 Procedure

The experiment was conducted online using Experigen (Becker & Levine, 2013). Before beginning, participants were required to listen to a real English word and type it correctly into a response box to ensure that they could hear the stimuli. They were then taken to a page with the instructions. Participants were told that they would be learning some words in a foreign language, and the procedure was explained to them.

The experiment itself had two phases: a training phase and a test phase. The training phase consisted of 64 trials (32 stems x 2 suffixes). Each trial began with an image of a singular object or animal in the center of the screen with a button below it. Clicking the button played the singular stem for the image. Presentation of the words was entirely auditory; participants never saw a written form for the words. A second image then appeared, which depicted either multiple of the object/animal (in plural trials) or a small version of it (in diminutive trials). Participants clicked two buttons, labeled "Option 1" and "Option 2", to hear the two suffixed options for the picture, one harmonic option and one disharmonic option (order counterbalanced). The second audio button only appeared after the first was clicked. After hearing both options, two response buttons appeared, again labeled "Option 1" and "Option 2"; participants chose the correct option by clicking one of the response buttons. Participants were forced to click each of the audio buttons sequentially (thus hearing all stimuli for the trial) before the response buttons appeared. Once participants clicked a response button, they received feedback printed on the screen ("Correct!" or "Incorrect. The correct response was Option 1/Option 2."). The audio buttons remained on the screen after appearing, and participants could listen to the words an unlimited number of times before advancing to the next trial.

The test phase consisted of 48 main trials (16 stems, 2 single suffix trials + 1 double suffix trial per stem) and 4 additional attention-check trials, randomly intermixed. The procedure of the single suffix trials was identical to training trials, except that feedback was no longer given. The procedure of the double suffix trials was similar, except that participants had four options instead of two (because each of the suffixes could independently be front or back). The diminutive suffix always occurred before the plural suffix in the response options for double suffix trials. Crucially, participants were not given any information about how to respond in double suffix trials.

Attention-check trials consisted of an initial word (e.g., *fast*) followed by four options (e.g., *deep, quick, blue, slow*); all were real English words spoken in a different (female) voice. Participants were explicitly told about the attention-check trials, and were instructed to choose the color word. In addition to the target color word, the response options included a synonym, an antonym, and an unrelated adjective. The goal was to provide multiple reasonable options so that participants who did not read the instructions carefully would fail the attention check.

After the test phase, participants completed a survey about their demographics, linguistic background, and response strategies. The full experiment lasted about 30 minutes.

2.2 Results and discussion

We first analyze the single suffix test trials, which were similar to the test trials reported in Pycha et al. (2003). We then analyze the double suffix trials, which were the crucial test trials in our design. The data were analyzed using mixed-effects logistic regression models implemented in R (Bates, 2014); all models contained random intercepts for Participants.

2.2.1 Single suffix test trials

Accuracy on single suffix trials is shown in Figure 1. Overall, we see very similar performance for participants in the harmony condition (59.7% correct) and participants in the disharmony condition (58.9% correct). We designed a model with the dependent variable, correct/incorrect, predicted by Condition (harmony or disharmony; contrast coded). We compared this model to a simpler model that excluded the predictor Condition using a likelihood ratio test. The simpler model was not significantly different from the full model ($\chi^2(1) < 1$), indicating that participants performed similarly in the harmony and disharmony conditions. To determine if participants showed learning in each condition, we ran a model with an intercept but no other fixed effects for each condition individually; these models showed that performance was above chance level in both conditions (harmony: $\beta = 1.00$, $SE = 0.22$, $\chi^2(1) = 18.7$, $p < 0.001$; disharmony: $\beta = 0.79$, $SE = 0.20$, $\chi^2(1) = 16.1$, $p < 0.001$). Thus, we see no evidence of differential learning of harmony and disharmony in single suffix test trials, replicating the findings of Pycha et al. (2003).

2.2.2 Double suffix test trials

Next, we turn to our analysis of the crucial double suffix test trials. Following our preregistered plan, this analysis consisted of two phases. The first analysis included all participants. For each double suffix test trial, the response was coded as correct if it was consistent with the general rule participants were hypothesized to learn, as in (1a) and (1b), and incorrect otherwise. Specifically, for the harmony condition, the response option where both suffixes were harmonic with the stem (schematically: FF-F-F or BB-B-B, where F = front vowel and B = back vowel) was coded as correct; for the disharmony condition,

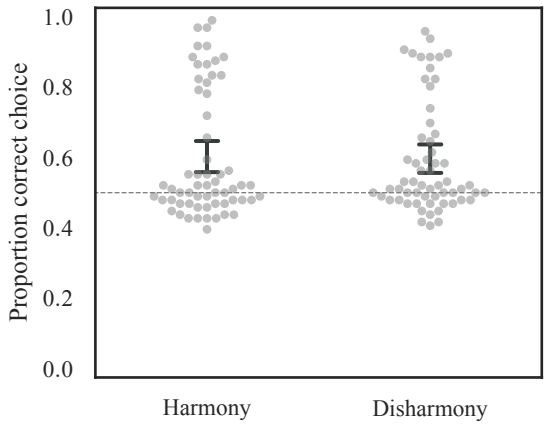

Figure 1: Proportion correct on single suffix trials by condition for all participants. Each point represents a participant and error bars represent 95% confidence intervals calculated on participant means. The broken line represents chance level (0.5).

the option with fully alternating backness values (schematically: FB-F-B or BF-B-F) was coded as correct. Figure 2 (left side) shows results aggregated by participant. On average, participants in the harmony condition chose the correct option 48.7% of the time, with chance level at 25% (as there were four response options). Participants in the disharmony condition chose the correct answer only 21.1% of the time, slightly below chance level.

We designed a mixed effects logistic regression model with the dependent variable, correct/incorrect, predicted by Condition (harmony or disharmony; contrast coded). We compared this model to a simpler model that excluded the predictor Condition using a likelihood ratio test. The simpler model was found to differ significantly from the full model ($\beta = 1.78$, $SE = 0.29$, $\chi^2(1) = 34.13$, $p < 0.0001$), indicating that participants in the harmony condition gave more correct responses than those in the disharmony condition.

Per our pre-registered analysis plan, we then conducted a restricted analysis including only those participants who demonstrated above-chance learning on single suffix test trials.

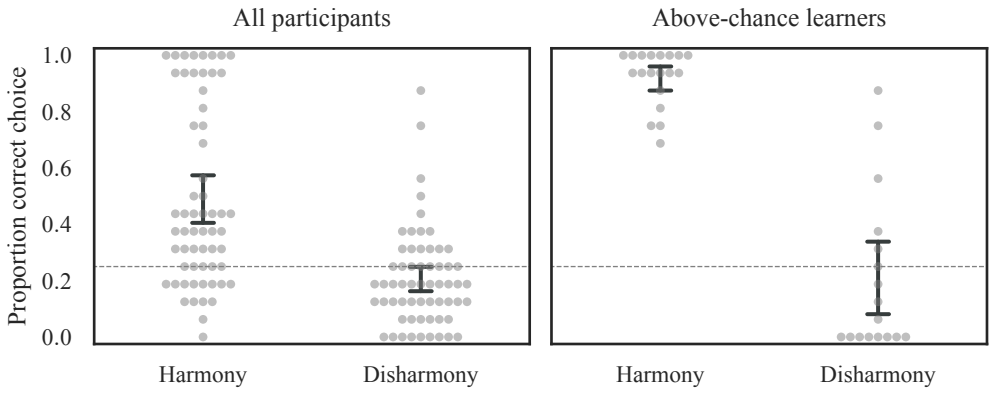

Figure 2: Proportion correct on double suffix trials by condition for all participants (left) and above-chance learners (right). Each point represents a participant and error bars represent 95% confidence intervals calculated on participant means. The broken line represents chance level (0.25).

Because single suffix test trials were identical to training trials, except with different stems, the secondary analysis includes participants who we can be confident robustly learned *some* pattern during training (i.e., only those individuals who are reliably above chance level in Figure 1). To be included in the secondary analysis, a participant must have responded correctly on 23/32 single suffix test trials, the minimum proportion that differs significantly from chance performance according to a two-tailed binomial test at an alpha level of 0.05. A total of 36 participants met this strict criterion (harmony condition: 19; disharmony condition: 17). Above-chance learners in the harmony condition overwhelmingly chose the correct option on double suffix test trials—92.1% of the time on average—whereas above-chance learners in the disharmony condition chose the correct response only 20.6% of the time, below the chance level of 25% (Figure 2, right). We ran the same statistical model detailed above on the subset of data from the above-chance learners. Again, the simpler model differed significantly from the full model ($\beta = 5.75$, $SE = 0.88$, $\chi^2(1) =$

$39.30, p < 0.0001$), indicating that participants in the harmony condition gave more correct responses than those in the disharmony condition.

These results reveal a striking distinction between harmony and disharmony. Participants in the harmony condition overwhelmingly inferred a pattern consistent with general harmony, especially for the above-chance learners.[4] By contrast, it is clear that participants in the disharmony condition did *not* learn a general disharmony pattern like (1b). Contrary to earlier findings (e.g., Pycha et al., 2003; Skoruppa & Peperkamp, 2011), our results show that there is a learnability difference between harmony and disharmony; that is, learners readily converge on an analysis consistent with general harmony but they do not converge on an analysis consistent with general disharmony. This distinction between harmony and disharmony only emerged when we required participants to extrapolate their learning to the unseen double suffix cases, a type of generalization not required in the earlier studies.

2.2.3 Analysis of response patterns

Participants in the disharmony condition clearly did not learn general disharmony, but what did they learn? We cannot be certain based on this experiment alone, but we can speculate by taking a closer look at the distribution of responses in the double suffix trials. We included in our pre-registration an additional (qualitative) exploratory analysis of the response patterns in both conditions. Recall that there were four response options in double suffix trials, three of which were coded as "incorrect" in the analysis above. The four possible response types are summarized in Table 1, with examples from the disharmony condition. Figure 3 shows the distribution across the four response options in each condition, for all participants (left) and above-chance learners (right). If participants responded

Description	Example	Label
Both suffixes harmonic w/ final stem V (and each other)	*peko-fu-bo*	HH
Suffix_1 disharm. w/ final stem V; suffix_2 disharm. w/ suffix_1	*peko-fi-bo*	DD
Suffix_1 disharm. w/ final stem V; suffix_2 harm. w/ suffix_1	*peko-fi-be*	DH
Suffix_1 harm. w/ final stem V; suffix_2 disharm. w/ suffix_1	*peko-fu-be*	HD

Table 1: Descriptions of the four possible response types on double suffix test trials (with examples from the disharmonic condition).

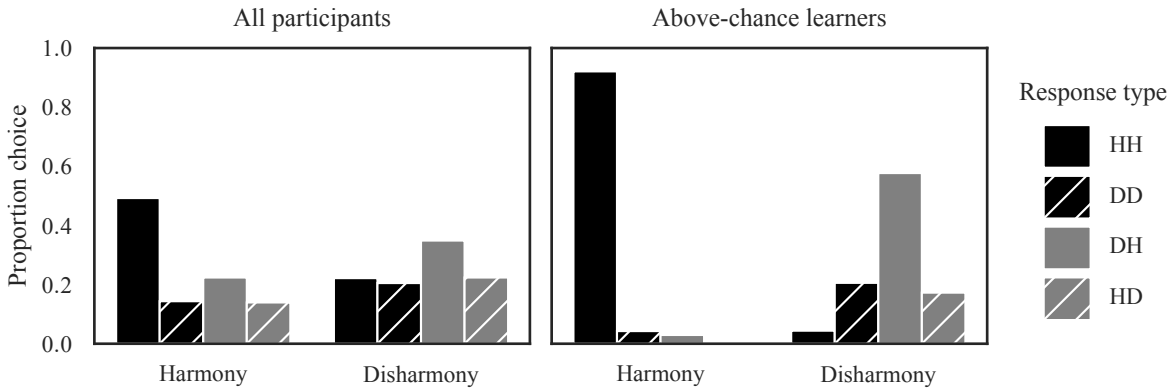

Figure 3: Proportion choice of each response type by condition, for all participants (left) and above-chance learners only (right).

completely at random, the bars in Figure 3 would all be around 0.25.

Looking first at the harmony condition, we see that the HH response (full harmony) was chosen far more often than any other response type, and nearly without exception by above-chance learners (echoing the results shown in Figure 2). In the disharmony condition, we find a much different pattern of responses. Participants showed no preference whatsoever for the DD response type, which is the correct response if they learned a general disharmony rule. Instead, they tended to prefer the DH option, meaning the first suffix was disharmonic with the final stem vowel, but the two suffixes were actually harmonic with each other.

There are at least two possible explanations for the pattern of responses seen in the

disharmony condition. Both explanations would account for why participants succeeded on single suffix trials but preferred the (incorrect) DH option in double suffix trials. First, participants may have focused on the stem-suffix morpheme boundary, learning that vowels must disagree in backness across the boundary. This interpretation would mean that participants in the disharmony condition tended to spontaneously infer *harmony* between the suffixes in untrained double suffix cases. This account would be consistent with the view that learners have a substantive bias favoring harmony. Such a bias would make it more likely for participants to infer a general harmony pattern when exposed to harmony (as in the harmony condition) and also more likely for them to assume harmony in the absence of evidence (as in the untrained double suffix cases in the disharmony condition).

A second possibility is that some participants learned specific correspondences between suffix allomorphs and stem types (e.g., *-fi* and *-be* combine with stems ending in *o*), and they concatenated the two allomorphs when responding to double suffix trials. Like the first possibility, this strategy produces the correct outcome for single suffix words, as well as the incorrect DH outcome preferred by many participants in double suffix words (e.g., *-fi* and *-be* are the correct disharmonic suffixes for stem *peko* individually, but *peko-fi-be* is incorrect). It is worth noting that in the harmony condition, this strategy would yield correct choices for both single suffix and double suffix trials, so we cannot rule out the possibility that some participants in the harmony condition also learned in this manner.

In fact, this possibility raises another potential type of learning pressure that could favor harmony over disharmony; that is, multiple analytical approaches to the data converge on a harmony analysis whereas only a specific (apparently not easily available) approach leads to a disharmony analysis. For example, let us assume some individual variation in

the approaches that learners take to analyzing input data. Some participants may prioritize generality and phonetic naturalness in the initial stages of learning, while others may rely more strongly on specific allomorph correspondences, as described above. With exposure to harmonic inputs, both strategies lead to a result where all vowels agree with their neighbors (which could in turn lead to grammaticalized vowel harmony). However, *only* the inference of a rule like (1b) yields a pattern where all vowels *dis*agree with their neighbors. This pressure could then play out at a population level, making general vowel disharmony less likely to be innovated and transmitted compared to harmony, given the narrower analytical approach needed for individual learners to arrive at the disharmony analysis.

Without further experiments, it is difficult to know for sure which strategy (or combination of strategies) participants in our task used. We did ask participants at the end of the experiment to describe their response strategies. Most answers were not explicit enough to be insightful. However, some participants in the harmony condition expressed strategies consistent with general harmony (e.g., "I noticed the vowels were always similar so I just went with that."), suggesting that at least some participants acquired a general harmony rule. By contrast, no participants in the disharmony condition expressed a strategy consistent with a general disharmony rule (e.g., something like "A vowel had to be different than the vowel next to it.").

3 Conclusions

The present study showed that learners do not readily infer a general vowel disharmony pattern, but they do converge on a general harmony pattern. The results indicate that, contrary to the conclusions of previous studies (e.g., Pycha et al., 2003; Skoruppa & Peperkamp, 2011), there is a fundamental learnability difference between vowel harmony and vowel disharmony patterns, which emerged when we required learners to extend what they learned, without additional training, from one suffix to two. One explanation for this distinction is a substantive bias favoring vowel harmony over disharmony, which would be in line with other recent work showing a role for substantive bias in phonological learning (e.g., Kimper, 2016; Martin, 2017; Martin & Peperkamp, forthcoming; Myers & Padgett, 2014; White, 2014). Another possibility is that a wider variety of analytical approaches converge on a harmony analysis compared to a disharmony analysis (in turn favoring harmony). These possibilities represent a rich field for future research, now that a clear asymmetry between vowel harmony and disharmony has been established.

Notes

[1]These approaches are not necessarily mutually exclusive (for discussion, see Beguš, 2018; Moreton, 2008; Moreton & Pater, 2012a, 2012b).

[2]Arguments have also been made for the perceptual motivation of vowel harmony, see Kaun (2004) and Kimper (2017), among others.

[3]Indeed, there is much experimental evidence that vowel harmony is learned better than similar (but formally more complex) arbitrary co-occurrence patterns (e.g., Baer-Henney

& van de Vijver, 2012; Baer-Henney, Kügler, & van de Vijver, 2015).

[4]It is worth considering whether our participants could have extrapolated a preference for harmony from their native language (English). A previous lexical analysis, however, found that the English lexicon contains relatively few harmonic words, no more than would be expected by chance (Martin & Peperkamp, forthcoming). Thus, our results cannot be explained by the lexical statistics of English.

References

Archangeli, D. & Pulleyblank, D. (1994). *Grounded Phonology*. Cambridge: MIT Press.

Baer-Henney, D., Kügler, F., & van de Vijver, R. (2015). The Interaction of Language-Specific and Universal Factors During the Acquisition of Morphophonemic Alternations With Exceptions. *Cognitive Science*, *39*(7), 1537–1569.

Baer-Henney, D. & van de Vijver, R. (2012). On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology*, *3*(2).

Bates, D. M. (2014). Lme4: Mixed-effects modeling with R.

Becker, M. & Levine, J. (2013). Experigen - an online experiment platform.

Beguš, G. (2018). *Bootstrapping sound changes*.

Blevins, J. (2004). *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Culbertson, J. & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*(16), 5842–5847.

Donegan, P. J. & Stampe, D. (1979). The study of natural phonology. In *Current Approaches to Phonological Theory* (pp. 126–173). Bloomington: Indiana University Press.

Finley, S. & Badecker, W. (2008). Analytic biases for vowel harmony languages. In *Proceedings of West Coast Conference on Formal Linguistics 27* (pp. 168–176).

Hayes, B. & Steriade, D. (2004). Introduction: the phonetic bases of phonological Markedness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 1–33). Cambridge: Cambridge University Press.

Kaun, A. (2004). The phonetic foundations of the rounding harmony typology. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 87–116). Cambridge: Cambridge University Press.

Kimper, W. (2016). Asymmetric generalisation of harmony triggers. In G. Ó. Hansson, A. Farris-Trimble, K. McMullin, & D. Pulleyblank (Eds.), *Proceedings of the 2015 Annual Meeting on Phonology*.

Kimper, W. (2017). Not crazy after all these years? Perceptual grounding for long-distance vowel harmony. *Laboratory Phonology: Journal for the Association of Laboratory Phonology*, *8*(1).

Maldonado, M. & Culbertson, J. (2019). Something about *us*: Learning first person pronoun systems. In A. K. Goel, C. M. Seifart, & C. Freksa (Eds.), *Proceedings of the 41st*

Annual Meeting of the Cognitive Science Society (pp. 749–755). Montreal, Canada: Cognitive Science Society.

Martin, A. (2017). *Biases in phonological processing and learning* (Doctoral dissertation, École normale supérieure).

Martin, A., Abels, K., Adger, D., & Culbertson, J. (2019). Do learners' word order preferences reflect hierarchical language structure? In A. K. Goel, C. M. Seifart, & C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (pp. 2303–2309). Montreal, Canada: Cognitive Science Society.

Martin, A. & Peperkamp, S. (forthcoming). Phonetically natural rules benefit from a learning bias: a re-examination of vowel harmony versus disharmony. *Phonology*.

Martin, A., Ratitamkul, T., Abels, K., Adger, D., & Culbertson, J. (2019). Cross-lingusitic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard*, *5*(1).

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*(01), 83–127.

Moreton, E. & Pater, J. (2012a). Structure and substance in artificial-phonology learning, part I: Structure. *Language and Linguistics Compass*, *6*(11), 686–701.

Moreton, E. & Pater, J. (2012b). Structure and substance in artificial-phonology learning, part II: Substance. *Language and Linguistics Compass*, *6*(11), 702–718.

Myers, S. & Padgett, J. (2014). Domain generalisation in artificial language learning. *Phonology*, *31*(3), 399–433.

Ohala, J. J. (1993). The phonetics of sound change. In *Historical linguistics: Problems and perspectives* (pp. 237–278).

Ohala, J. J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. In *Proceedings of the 3rd Inernational Conference on Spoken Language Processing*.

Pycha, A., Nowak, P., & Shin, E. (2003). Phonological Rule-Learning and Its Implications for a Theory of Vowel Harmony. In *Proceedings of WCCFL* (Vol. 22, pp. 423–435).

Rose, S. & Walker, R. (2011). Harmony systems. In *The Handbook of Phonological Theory* (pp. 240–290).

Skoruppa, K. & Peperkamp, S. (2011). Adaptation to Novel Accents: Feature-Based Learning of Context-Sensitive Phonological Regularities. *Cognitive Science*, *35*(2), 348–366.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*(1), 96–115.

White, J., Kager, R., Linzen, T., Markopoulos, G., Martin, A., Nevins, A., … van de Vijver, R. (2018). Preference for locality is affected by the prefix/suffix asymmetry: Evidence from artificial language learning. In *Proceedings of NELS 48*.

Wilson, C. (2006). Learning Phonology With Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science*, *30*(5), 945–982.