## Coarticulation and learnability of transparent vowels in vowel harmony

Sara Finley*


#### Abstract

The representations of transparent vowels in vowel harmony have been of interest to phonologists because of the challenges they pose for constraints on locality and complexity. One proposal is that transparent vowels in back vowel harmony may be intermediate between front and back. The present study uses two artificial language learning experiments to explore the psychological reality of acoustic differences in transparent vowels in back vs. front vowel contexts. Participants were exposed to a back/round vowel harmony language with a neutral vowel that was spliced so that the F2 was lower in back vowel contexts and higher in front vowel contexts (the Natural condition) or the reverse (the Unnatural condition). While only participants in the Natural condition of Experiment 1 were able to learn the behavior of the transparent vowel relative to a No-Training control, there was no difference between the Natural and Unnatural conditions. In Experiment 2, only participants in the Natural condition learned the vowel harmony pattern, though there were no significant differences between the two conditions. No condition successfully learned the behavior of the transparent vowel in Experiment 2. These results suggest that the effects of small differences in the F2 value of transparent back vowels on learnability are minimal.


Keywords. vowel harmony; transparent vowels; learnability; coarticulation; artificial language learning; phonology

1. Introduction. Vowel harmony is one of the most commonly studied phenomena in phonology, primarily because it poses questions about representation and action at a distance. Vowel harmony is characterized as a non-local process for a variety of reasons. First, vowel harmony typically applies only to vowels, skipping consonants (though vowel harmony can interact with consonants, Turkish being a commonly cited example (Clements \& Sezer 1982)). Second, vowel harmony often applies iteratively, applying to multiple segments, rather than just adjacent vowels. For example, in Turkish, an agglutinative language, vowel harmony applies to all suffixes, not just the one closest to the stem (e.g., [ip-ler-in] 'rope, GenPl', and [son-lar-in] 'end, GenPl') (Clements \& Sezer 1982). Third, it is possible for neutral/non-participating vowels to be transparent to vowel harmony. In the case of transparent vowels, harmony 'skips' over neutral vowels, thereby adding an additional level of non-locality. For example, in Hungarian, the backness of the dative suffix in [radir-nak] 'easier-dat' is determined by the initial stem vowel, and skips over the transparent front vowel (Ringen 1988).

The question of how to account for the locality issues of transparent vowels has been a longstanding concern in phonological theory. At issue is how to account for the non-local components of transparency, while still maintaining locality effects that apply generally in languages. For example, many approaches to vowel harmony in Optimality Theory (Prince \& Smolensky

[^0]2004) either suffer from over-predicting non-local effects (Wilson 2003a), or are unable to account for transparent vowels without additional machinery, such as targeted constraints (Baković \& Wilson 2000) or serial order (Kimper 2011).

Many representational proposals for transparent vowels in vowel harmony are supported by the phonetic realization of transparent vowels (Finley 2008; Smith 2016; Rebrus \& Törkenczy 2015). For example, Finley (2008) uses a modified version of turbidity theory (Goldrick 2000), where intermediate representations allow for transparent vowels to undergo spreading at an abstract level, but not on the surface level. Rebrus and Törkenczy (2015) propose that transparent vowels in back harmony are intermediate between front and back. Smith (2016) uses a gestural approach, where the gestural representation of transparent vowels is different from opaque vowels. The gestural approach is based on the articulatory properties of transparent vowels in vowel harmony. Benus and Gafos (2007) showed that Hungarian speakers produce transparent vowels differently depending on their harmonic contexts. Despite being phonologically front, transparent vowels are produced more retracted in back vowel contexts compared to front vowel contexts. Gick et al. (2006) showed that transparent, [-ATR] low vowels in Kinande are categorically produced as [+ATR] in [+ATR] contexts, despite the fact that low vowels are not contrastive for ATR, suggesting a phonetic implementation of transparency that goes beyond mere coarticulation (as the effect did not diminish with distance from the trigger).

The fact that many phonological accounts of transparent vowels are inspired by the phonetics of transparent vowels supports the hypothesis that phonetic factors that relate to vowel harmony, even if subtle, may be at play in the psychological representations of transparent vowels. In addition, such factors may play a role in the learnability of vowel harmony patterns. Previous research using the artificial grammar learning paradigm has suggested that adult learners may be sensitive to the phonetic factors that give rise to typological constraints on vowel harmony (Finley \& Badecker 2012; Finley \& Badecker 2009; Kimper 2016; Finley 2012; Pycha et al. 2003; Martin \& White 2021). For example, participants were more likely to learn and generalize a round vowel harmony language that had non-high vowel triggers (Kimper 2016; Finley 2012), which are typologically more likely to trigger harmony (Kaun 2004). This suggests that naïve learners of a vowel harmony language with transparent vowels might benefit from stimuli that enhances phonetic properties of transparent vowels, compared to a language that does not.

However, there is some reason to be skeptical that the articulatory differences between transparent vowels in various contexts may have any effect on learning and generalization. Some research on the acoustics of transparent vowels in Hungarian showed null or mixed results (Blaho \& Szeredi 2013; Szeredi 2016). While the acoustic differences trended in the direction shown in Benus and Gafos (2007), they did not always reach statistical significance, suggesting that the effects on perception and learnability might be minimal. While Szeredi (2016) found small acoustic differences between vowels in different contexts, the Just Noticeable Difference (JND) for Hungarian speakers was between 200 and 250 Hz , much larger than the measured differences. This suggests that even if the effects are in the acoustic signal, they may not be a part of speakers' active, cognitive representation of Hungarian vowel harmony.

In addition, while there have been several studies that showed learning biases towards phonetically natural patterns in artificial language learning settings (Carpenter 2010; Finley 2012; Pycha et al. 2003; Wilson 2003b; Martin \& White 2020), other studies have shown more mixed results (Glewwe 2019; Wilson 2006; Skoruppa \& Peperkamp 2011), and the existence of phonet-ically-based restrictions on the learnability of phonological patterns has been questioned (Moreton \& Pater 2012b). Rather than phonetic naturalness affecting learning ability, abstract
constraints on cognition and linguistic representations may shape biases in phonological learning (Moreton \& Pater 2012a). Artificial language learning studies that explored learning biases from a more cognitive or representational point of view have shown more consistent patterns of results (Moreton, Pater \& Pertsova 2017; Finley 2017; Do 2018; Burness \& McMullin 2019).

The present study serves to test the hypothesis that a vowel harmony pattern with transparent vowels that follow more natural constraints on coarticulation should be easier to learn than a harmony pattern where transparent vowels show unnatural coarticulatory effects. Participants were exposed to vowel harmony with a transparent vowel that was cross-spliced from either a front or back vowel context in natural speech, but then spliced into stimuli in either a natural or an unnatural manner. In the Natural condition, front vowels were further front in front vowel contexts, and in the Unnatural condition, front vowels were further back in front vowel contexts. If learners are sensitive to these differences, then participants in the Natural condition, where coarticulation is enhanced, should be more likely to learn the behavior of the transparent vowel than in the Unnatural condition.
2. Experiment 1. Experiment 1 tested for the role of coarticulation in learning transparent vowels in vowel harmony.
2.1. Participants. There were 57 ( $\mathrm{n}=19$ in each condition) adult, English speaking participants whose data were included in the experiment. Participants were recruited from the subject pool of the psychology department at a small liberal arts college in the Pacific Northwest of the United States. Course credit was given for participation. No participant reported any known exposure to a vowel harmony language, natural or artificial.
2.2. DESIGN AND MATERIALS. Participants in the Experimental conditions were exposed to an artificial language with a back/round vowel harmony pattern with a transparent vowel. Much of the stimuli were identical to those used in Finley (2015), where the harmony pattern was exemplified through stem, stem+suffix alternations (e.g., /bidig, bidige/). Stems were always of the form CVCVC, and the suffix was always either /e/ or /o/, depending on the harmonic features of the stem, which always adhered to back/round harmony with a transparent vowel. Consonants were chosen from the set $/ \mathrm{p}, \mathrm{t}, \mathrm{k}, \mathrm{b}, \mathrm{d}, \mathrm{g}, \mathrm{m}, \mathrm{n} /$, and vowels were chosen from the set $/ \mathrm{e}, \mathrm{i}, \mathrm{o}, \mathrm{u}, \mathrm{I} /$ where $/ \mathrm{I} /$ was the transparent vowel. There were 24 total pairs of stem, stem + suffix items, six each from the following categories: front vowel stems, where both vowels were front (and always took the front vowel [e]), back vowel stems, where both vowels were back (and always took the back vowel $[\mathrm{o}]$ ), transparent front vowel stems, where the initial vowel was front, and the second vowel was the transparent vowel (and always took the front vowel [e]), transparent back vowel stems, where the initial vowel was back, and the second vowel was the transparent vowel (and always took the front vowel [o]). Examples ${ }^{1}$ of training items can be found in Table 1.

|  | Stem | Stem+Suffix |
| :--- | :--- | :--- |
| Front Vowel $(\mathrm{n}=6)$ | pikeg | pikege |
| Back Vowel $(\mathrm{n}=6)$ | boduk | boduko |
| Transparent Front Vowel $(\mathrm{n}=6)$ | gimit | gimıte |
| Transparent Back Vowel $(\mathrm{n}=6)$ | domıd | domıdo |

Table 1. Examples of training items

[^1]The difference between the Natural and the Unnatural conditions were based on the quality of the transparent vowel. While the vowel was always $/ \mathrm{I} /$, the $/ \mathrm{I} /$ was spliced into each sound file from two tokens, one from a front vowel context ([dimike]) with an average F2 of 2492 Hz , and one from a back vowel context ([pomiko]), with an average F2 of 2212 Hz . In the Natural condition, the transparent vowel [I] was always in a 'correct' or natural context: the high F2 [r] in front vowel contexts, and the lower F2 in back vowel contexts. In the Unnatural condition, the [I] was always spliced into an 'incorrect' or unnatural context: the high F2 [r] in back vowel contexts, and the low F2 [I] in front vowel contexts [I]. All splicing was conducted using Praat software (Boersma \& Weenink 2017).

All stimuli were naturally-produced, recorded in a sound-attenuated booth by an adult, female, native English speaker. The speaker had no knowledge of the specific hypothesis of the study, but was aware that that stimuli were to be used in an artificial language learning experiment. The talker repeated each word four times, and one of the two tokens in the middle (second and third) was chosen to be used for the experiment. The middle items were chosen in order to avoid list-like prosody in the items. The talker was told to produce all vowels, but primary stress was placed on the final syllable. All stimuli were normalized to 70db in Praat (Boersma \& Weenink 2017), but participants were allowed to adjust the volume to their needs.

Learning of the harmony pattern and the behavior of the transparent vowel was assessed with a two-alternative forced choice task, where participants were asked to select between two suffixed forms with the same stem; one choice ended in [e] and the other ended in [o]. There were four different types of test items, with examples in Table 2.

|  | Front Vowel Suffix | Back Vowel Suffix |
| :--- | :--- | :--- |
| Old (n=10) (Experiment 1 only) | pikege | *pikego |
|  | *bupite | bupito |
| New $(\mathrm{n}=10)$ | *pokıge | pokıgo |
|  | *nupuke | nupuko |
| New-Transparent Front $(\mathrm{n}=10)$ | tenine | *tenino |
| New-Transparent Back $(\mathrm{n}=10)$ | degibe | *degibo |
| New-Transparent Back Unspliced $(\mathrm{n}=9)$ | *todıpe | todipo |
| (Experiment 2 only) |  | todipo |

Table 2. Examples of test items (*indicates incorrect response)
Old items were items that were part of the training set, and included both items with the transparent vowel, and items without it. New items included items that were not in the training set, but did not have the transparent vowel. New-Transparent Front items included items where the initial vowel was front, and the second stem vowel was the transparent vowel. New-Transparent Back items included items where the initial vowel was back, and the second stem vowel was back. The New-Transparent items consisted of the same splicing that occurred in the training set. The Old items contained a mix of items with and without the transparent vowel, and were thus slightly different between the Natural and the Unnatural condition. Because the New items contained no splicing, these items were identical for both conditions. Examples of test items can be found in Table 2. The No Training Control condition responded to all sets of items (e.g., both sets of New-Transparent Front).
2.3. Procedure. The experiment was run using PsyScope software (Cohen et al. 1993), in a standard laboratory setting. Participants in the experimental condition were told that they would
be learning a novel language, and that their task was to listen to the words, presented as pairs over headphones. Each training trial consisted of a pair of items (stem followed by the stem+suffix form), repeated 10 times, each in a random order. Immediately following the training phase, participants completed the two alternative forced choice test. Participants in the experimental conditions were told that they would be hearing two words, and that their job was to select the word that was most likely to belong to the language that they had just heard. If the first word was more likely to belong to the language, they were to press the ' $a$ ' key, and if the second word was more likely to belong to the language, they were to press the ' $l$ ' key. The order of each option was the same for all participants, but balanced between whether the first option was correct, and whether the second option was correct, as well as whether the first option contained front vowel stems or back vowel stems (where possible). Participants in the no training Control condition were only given test items, and were thus told to select the word that they preferred based on any criteria they chose. All participants were debriefed upon completion of the experiment. The experiment took approximately 20 minutes to complete for the Experimental conditions, and less than 10 minutes to complete for the Control condition.
2.4. ReSults. Means (and standard errors of the mean) for Experiment 1 can be found in Figure 1. There is only a single bar representing the Control conditions for New items because these items (and results) were identical. The data were fit into generalized linear mixed effect models fit by the Laplace approximation using the lme4 (Bates et al. 2015) package in $\mathrm{R}(\mathrm{R}$ Development Core Team 2018) via R Studio (RStudio Team 2020). ${ }^{2}$


Figure 1. Experiment 1 Results (Means and Standard Errors of the Mean)
Due to convergence errors, the Old items were removed from analysis, and three separate models were run with the New-Transparent Back items set as the baseline. In general, participants in both conditions performed well on Old items (mean $=0.77 \pm 0.060$ for the Natural condition; mean $=0.72 \pm 0.063$ for the Unnatural condition), suggesting that participants were able to remember the items that appeared in the training set. One model compared the Natural to its No-

[^2]Training control counterpart (Table 3), a second model compared the Unnatural condition to its No-Training Control counterpart (Table 4), and a third model compared the Natural and the Unnatural conditions to each other (even though these were technically different items due to the differences in splicing) (Table 5).

|  | $\beta$ | SE | z | p |
| :--- | :--- | :--- | :--- | :--- |
| Intercept | 0.36 | 0.21 | 1.71 | 0.087 |
| Natural.Control | -0.58 | 0.29 | -1.97 | 0.048 |
| NewT-Back.New | 0.24 | 0.22 | 1.11 | 0.27 |
| NewT-Back.New-T-Front | 0.50 | 0.22 | 2.23 | 0.026 |
| Natural.Control: NewT-Back.New-T-Front | 0.17 | 0.30 | 0.56 | 0.57 |
| Natural.Control: NewT-Back.New | -0.11 | 0.31 | -0.35 | 0.73 |

Table 3. Model with New-Transparent Back, Natural as Baseline
Participants in the Natural condition showed significantly more correct responses to NewTransparent Back vowel items compared to the Control condition. There were significantly more correct responses to New-Transparent Front vowel items compared to New-Transparent Back vowel items, but no differences between New and New-Transparent Back vowel items, and no interactions by Condition and test item.

|  | $\beta$ | SE | z | p |
| :--- | :--- | :--- | :--- | :--- |
| Intercept | 0.36 | 0.19 | 1.92 | 0.055 |
| Unatural.Control | -0.32 | 0.27 | -1.19 | 0.23 |
| NewT-Back.New | 0.52 | 0.22 | 2.37 | 0.018 |
| NewT-Back.New-T-Front | 0.49 | 0.22 | 2.56 | 0.024 |
| Unnatural.Control: NewT-Back.New-T-Front | -0.34 | 0.30 | -1.13 | 0.26 |
| Unnatural.Control: NewT-Back.New | -0.36 | 0.30 | -1.19 | 0.23 |

Table 4. Model with New-Transparent Back, Unnatural as Baseline
Participants in the Unnatural condition showed no significant differences between the Control condition for New-Transparent Back items. There were significantly more correct responses to both New and New-Transparent Front items compared to New-Transparent Back items, but no interactions by Condition and test item. There were no significant differences between the Natural and Unnatural condition for New-Transparent Back items, and there were no interactions by Condition and test item.

|  | $\beta$ | SE | z | p |
| :--- | :--- | :--- | :--- | :--- |
| Intercept | 0.45 | 0.31 | 1.47 | 0.14 |
| Natural.Unnatural | -0.036 | 0.40 | -0.091 | 0.93 |
| Natural.Unnatural:NewT-Back.New-T-Front | 0.28 | 0.32 | 0.89 | 0.37 |
| Natural.Unnatural:NewT-Back.New | -0.013 | 0.32 | -0.041 | 0.97 |

Table 5. Model with New-Transparent Back, Unnatural as Baseline
Participants in Experiment 1 appear to have learned the general vowel harmony pattern in both the Natural and the Unnatural conditions. However, only participants in the Natural condition appeared to show learning of the behavior of the transparent vowel in back vowel contexts when compared to its No-Training Control counterpart.
2.5. Discussion. Because there was no significant difference between the Natural and the Unnatural conditions for New-Transparent Back vowel items, it is unclear how to interpret the fact that only participants in the Natural condition were significantly different from their control counterpart for New-Transparent Back items. One possibility is that idiosyncrasies in the items may have led to more variable responses in the Control condition, which led to differences in which conditions showed statistical significance. Since the New-Transparent items contained the splicing from the training set, the participants in the Natural and Unnatural conditions responded to slightly different items, making direct comparison of the two conditions difficult. Experiment 2 remedies this by including a set of New-Transparent items without splicing in order to allow for more direct comparison of the two conditions.
3. Experiment 2. Experiment 2 replicated Experiment 1, but included an additional set of test items in order to directly compare the Natural and Unnatural conditions.
3.1. Participants. There were 45 adult, English speaking participants whose data were included in the experiment ( $\mathrm{n}=23$ for the Natural condition, and $\mathrm{n}=22$ for the Unnatural condition). Participants were recruited from the same subject pool Experiment 1. Participants had no known exposure to a vowel harmony language, natural or artificial, nor did they participate in Experiment 1. Participants were given course credit for their participation.
3.2. Design and Materials. Experiment 2 was identical to Experiment 1, except for the following changes. First, the Control condition was removed in order to better allow for comparisons between the Natural and the Unnatural conditions. Second, the test included an additional set of Unspliced New-Transparent Back items. Because of access to original, unspliced stimuli was limited, there were only nine Unspliced items. Third, because Old items were not compared in statistical analyses in Experiment 1, there were no Old items in Experiment 2.
3.3. Procedure. The procedure for Experiment 2 was identical to Experiment 1, except that 14 participants ( $\mathrm{n}=7$ in each condition) were run outside of the laboratory, using FindingFive, a web-based platform for collecting data online (FindingFive Team 2019) (due to the COVID-19 pandemic). The online procedure was similar to the in-laboratory procedure, except for some minor details. First, participants were asked to use their own headphones, and complete the task in one sitting in a quiet, relatively isolated location, in order to better emulate a laboratory setting. Participants were given a sample audio file (/udvu/) to listen to as an opportunity to check the sound quality at their location. Participants were asked not to proceed with the study until their internet connection and sound quality were sufficient. Second, each training trial required participants to press a 'Continue' button in order to advance to the next trial. This was done instead of auto-advancement of trials in order to ensure that the participants were engaged in the task for the duration of the training (preventing the participant from leaving the computer as the training trials completed). Third, participants were encouraged to take a short break between the training and the test phases. Finally, participants were given a short post-completion survey following debriefing. This survey allowed participants to provide feedback (if any), and a chance to recuse themselves from inclusion in the final data analysis if they did not wish their data to be used. This was done to give participants who were not engaged in the task a way to earn credit for their participation, but not contaminate the results; all 14 participants agreed to have their data used in the analysis.
3.4. Results Means (and standard errors of the mean) for Experiment 2 can be found in Figure
2. The data were analyzed in the same manner as Experiment 1, except because there was no

Control condition, comparisons were only made between the two critical conditions. In addition, trials from the online participants were dropped if the participant took longer than 10s to respond ( $\mathrm{n}=20$ ), and an additional trial was dropped from an in-lab participant because the participant did not follow directions on the first trial, and required a repetition of the directions in between hearing the first item and responding to that item.


Figure 2. Experiment 2 Results (Means and Standard Errors of the Mean)
As shown in Table 6 and Table 7, there were no significant differences between the Natural and the Unnatural conditions for the Unspliced items for New-Transparent Back vowel items. In addition, the intercepts for both models were not significant, suggesting that the participants in Experiment 2 did not select the correct responses to New-Transparent Back vowel items at a rate greater than chance.

|  | $\beta$ | SE | z | p |
| :--- | :--- | :--- | :--- | :--- |
| Intercept | -0.042 | 0.20 | -0.21 | 0.83 |
| Unatural.Natural | 0.32 | 0.26 | 1.22 | 0.22 |
| Unspliced.New-T-Back | -0.10 | 0.22 | -0.46 | 0.64 |
| Unspliced.New-T-Front | 0.37 | 0.22 | 1.68 | 0.094 |
| Unspliced.New | 0.087 | 0.23 | 0.38 | 0.70 |
| Unnatural.Natural:Unspliced.New-T-Back | 0.092 | 0.29 | 0.32 | 0.75 |
| Unnatural.Natural:Unspliced.New-T-Front | -0.14 | 0.29 | -0.49 | 0.62 |
| Unnatural.Natural:Unspliced.New | 0.015 | 0.29 | 0.053 | 0.96 |

Table 6. Model with Unspliced Items, Unnatural Condition as Baseline

In the Unnatural condition, there were no significant differences between the Unspliced items, and any of the other types of test items, including New-Transparent Back items, New-Transparent Front, or New items, suggesting overall poor performance on all item types.

|  | $\beta$ | SE | z | p |
| :--- | :--- | :--- | :--- | :--- |
| Intercept | 0.27 | 0.20 | 1.40 | 0.16 |
| Unspliced.New-T-Back | -0.011 | 0.22 | -0.048 | 0.96 |
| Unspliced.New-T-Front | 0.23 | 0.23 | 1.02 | 0.31 |
| Unspliced.New | 0.10 | 0.23 | 0.45 | 0.65 |

Table 7. Model with Unspliced, Natural as Baseline
In the Natural condition, there were no significant differences between the Unspliced items, and any of the other types of test items, including New-Transparent Back items, New-Transparent Front, or New items. There were also no interactions between the Natural and the Unnatural condition in terms of the differences between the Unspliced items and the other items.

Because participants in neither the Natural nor the Unnatural conditions showed evidence of learning in the Unspliced (via significant intercepts), and with no significant differences between the Unspliced items and the other conditions, additional models were run to test whether the intercepts were significantly above chance for the overall experiment. There was a significant intercept for the Natural condition (mean $=0.58, \pm 0.033$ ), $\beta=0.36, \mathrm{SE}=0.14, \mathrm{z}=2.63, p$ $=0.0085$, suggesting overall learning of the harmony pattern. The intercept for the Unnatural condition (mean $=0.51, \pm 0.034$ ) was not significant, $\beta=0.052$, $\mathrm{SE}=0.14, \mathrm{z}=0.38, p=0.70$, suggesting that participants in the Unnatural condition did not learn the overall harmony pattern. The overall difference between the Natural and the Unnatural condition was only marginally significant $\beta=0.31, \mathrm{SE}=0.19, \mathrm{z}=1.65, p=0.099$.

In order to more fully investigate whether the failure to learn the overall harmony pattern in the Unnatural condition was due to the low responses for the Unspliced items, additional, post hoc models were run on the New, New-Transparent Back, and New-Transparent Front items. There was a significant intercept for New items for the Natural condition, $\beta=0.38, \mathrm{SE}=0.19, \mathrm{z}$ $=1.98, p=0.048$, but not the Unnatural condition, $\beta=0.045, \mathrm{SE}=0.19, \mathrm{z}=0.23, p=0.82$, though the difference between the two conditions was not significant, $\beta=0.33, \mathrm{SE}=0.25, \mathrm{z}=$ $1.32, p=0.19$. There was a significant intercept for the New-Transparent Front items in the Natural condition, $\beta=0.51, \mathrm{SE}=0.19, \mathrm{z}=2.65, p=0.0081$, but only a marginally significant intercept in the Unnatural condition, $\beta=0.33, \mathrm{SE}=0.19, \mathrm{z}=1.77, p=0.077$, but these were not significantly different from each other, $\beta=0.18, \mathrm{SE}=0.25, \mathrm{z}=0.69, p=0.49$. There was no significant intercept for the New-Transparent Back items in either the Natural condition, $\beta=0.26$, $\mathrm{SE}=0.19, \mathrm{z}=1.40, p=0.16$, or the Unnatural condition, $\beta=-0.15, \mathrm{SE}=0.19, \mathrm{z}=-0.78, p=$ 0.44 , and these were not significantly different from each other, $\beta=0.41, \mathrm{SE}=0.25, \mathrm{z}=1.62, p$ $=0.11$. These results suggest that overall, participants in Experiment 2 struggled to learn the basic harmony pattern as well as the behavior of the transparent vowel. While this struggle was more apparent in the Unnatural condition, there were no significant differences between the two conditions.
3.5. DISCUSSION. Experiment 2 was a replication of Experiment 1, with the addition of Unspliced items, which were unspliced versions of New-Transparent Back test items used in both experiments 1 and 2. The addition of these test items allowed for a more direct comparison between the two conditions, rather than through comparisons between different Control conditions. Participants in the Natural condition were able to learn the general harmony pattern, but did not show any evidence of learning the behavior of the transparent vowel in back vowel contexts. Participants in the Unnatural condition showed no evidence of learning the vowel harmony language. These differences should be taken with caution, because comparisons between the Natural and the Unnatural condition were generally not statistically significant.

Given that the training in Experiment 2 was identical to that of Experiment 1, there is question of why learning was so weak in Experiment 2. Even the Natural condition, which showed evidence of learning, still had an overall mean of less than $60 \%$ correct. One possibility is that several of the participants had to be run online. Because the acoustic differences between the Natural and Unnatural conditions are relatively subtle, it is possible that the participants in Experiment 2 were not well-equipped to hear the materials properly outside a controlled laboratory setting. However, the overall means in the two modalities were very similar. The overall means for the Natural condition run in the lab were both 0.58 , and the overall means for the Unnatural conditions was slightly higher for those who were run online ( 0.53 online compared to 0.50 in the lab). Thus, it is unlikely that the setting for the experiment played a major role in the differences between experiments.

Another possibility is that the lower performance in Experiment 2 could have been an artifact of the timing of data collection for Experiment 2, which occurred at the end of the semester, over the course of several semesters. These two factors may have affected the population of student sampled, particularly participant motivation. Participants in Experiment 2 were generally run at the end of the semester, when participants are seeking extra credit to alleviate poor grades, are scrambling to finish their research requirement, or are just feeling exhausted and unmotivated from the stress of the academic term. Participants in Experiment 1 were run throughout the semester, thus avoiding possible declines in motivation and attention, though the evidence of a significant end-of-semester effect is mixed (Ebersole et al. 2016; Grimm, Markman \& Maddox 2012; Nicholls et al. 2015). It is also possible that demographic shifts in the student population may have varied, such that the participants were generally less engaged in the experimental task over time, since Experiment 1 was run about two to three years before Experiment 2.

A third possibility for the overall lower rates of correct responses in Experiment 2 is that the test items themselves may have made inadvertently shifted participants' responses. In Experiment 2, there were essentially two sets of New-Transparent Back items, which were almost identical except that one contained splicing and the other did not. Having two sets of similar items in a two-alternative forced choice test may have influenced the participants' responses. Since a two-alternative forced choice test shows both the correct and the incorrect response, hearing the incorrect response twice for the same set of items may have reduced learners' confidence in their responses. However, a separate, post-hoc model with trial number as a fixed effect was not significant, $\beta=0.0020, \mathrm{SE}=0.0033, \mathrm{z}=0.61, p=0.54$. The first half of trials showed a mean of $57 \%$ correct responses, and the second half of trials showed a mean of $54 \%$ correct. Thus, in order for the repeated trials to have had a significant effect on responses it would have had to show up very early in the test.

While the Natural condition was the only condition in Experiment 2 to successfully learn the vowel harmony language, it is not clear that this advantage was entirely due to the splicing, rather than a statistical artifact, as the differences between the Natural and the Unnatural conditions were not statistically significant. Thus, it is unclear what, if any advantage coarticulation may have on learning vowel harmony with transparent vowels.
4. General discussion. I have presented the results from two artificial language learning experiments on vowel har-mony with a transparent vowel. The transparent vowel was spliced to either enhance coarticulatory effects (the Natural condition), or opposite coarticulatory effects (the Unnatural condition). In Experiment 1, participants in the Natural condition showed significantly more correct responses in the critical trials compared to a no-training control condition, but this was not shown for the Unnatural condition. However, there were no significant differences between the two experimental conditions. In Experiment 2, participants were given the same training as in Experiment 1, with slightly different test items, specifically with a set of unspliced critical transparent vowel items (in order to better compare the Natural and the Unnatural conditions). In Experiment 2, neither condition showed evidence of learning the behavior of the transparent vowel in critical back vowel contexts, and only the Natural condition showed evidence of learning the vowel harmony pattern.

While the results of the two experiments generally favor the Natural condition, they are not statistically robust enough to make any strong conclusions regarding the role of coarticulation in learning transparent vowels. There are several possible explanations for these results. First, the effect sizes may have simply been too small to detect with the current sample sizes. This may be especially true given that many participants may not have been motivated to learn the harmony pattern, which may have added additional noise to the experiment. A replication with much larger sample sizes could tease this apart. A replication that included catch trials to rule out participants who are not attending to the task could help to better understand the effects in the present study, such as explicit vs. implicit learning strategies (Moreton \& Pertsova 2017). A fol-low-up study could also include adjusting the training so that participants are only trained on the transparent vowel once they have successfully learned the regular harmony pattern. This could make it easier to interpret failure to learn the behavior of the transparent vowel if all participants have learned the regular harmony pattern. In Experiment 1, only three participants (two in the Natural, and one in the Unnatural conditions) scored less than $60 \%$ on Old items, suggesting that most participants were able to remember the items, and were able to follow directions and attend to the task. In Experiment 2, enough participants failed to learn the harmony pattern that comparison between learners and non-learners was not informative.

Second, the differences between the Natural and the Unnatural condition were relatively small and subtle. Szeredi (2016) tested the perceptibility of Hungarian transparent vowels on native Hungarian speakers, and found that the perceptibility of the vowels was between 200 and 250 Hz . While the stimuli used in the present experiment is in this range, it is unclear whether participants, who have no prior exposure to these differences, were able to both perceive the differences, and use those differences in learning. One possibility for future research could be to
transparent vowels in different contexts, or directly within Praat, the method used by Szeredi (2016). However, it is unclear what kind of external validity such effects would have, as the acoustical measurements for Transparent vowels in Hungarian were quite small, and showed many idiosyncrasies, including some participants showing an effect in the opposite direction (Szeredi 2016; Blaho \& Szeredi 2013).

If the coarticulation effects on transparent vowels applies only in articulation, as suggested in previous research (Blaho \& Szeredi 2013; Szeredi 2016), then the effects on learning might only be captured using a production-based task. Glewwe (2019) suggests that substantive biases in phonology arise from production, rather than perception, and production-based learning tasks are the best, if not only way to capture these effects. It is therefore possible that if the present study had used a production task instead of a perception-based task, that participants might have shown a significant effect ${ }^{3}$ in the hypothesized direction.

Third, learning the behavior of transparent vowels in an artificial language learning task with English speakers has been shown to be a challenge. Finley (2015) showed that learners are biased towards opaque neutral vowels compared to transparent neutral vowels. While the present study used almost identical training and test items to the conditions that allowed for learning transparent vowels, the introduction of the spliced items may have impeded learning, since crosssplicing reduced the naturalness of the stimuli (even though spoken).

Fourth, the present study made use of a back/round vowel harmony pattern, but tested coarticulation for backness only (i.e., F2). It is therefore possible that the phonetic constraints on transparency may be different for back/round harmony. Participants were given back/round harmony largely because most of the vowels in the English phonetic inventory conflate backness and rounding, making it impossible to replicate back vowel harmony for English speakers without using vowels that are not found in English. A follow-up study could test speakers with a wider vowel inventory (e.g., French or German) on backness harmony without rounding. In addition, more research on the phonetics of neutral vowels could be done on a wider range of languages, including back/round harmony.

Finally, a possible reason for the lack of a clear result is that coarticulation does not affect the learnability of neutral vowels in an artificial language learning setting with adults. If there is truly no effect, then no manipulation of the coarticulation or sample size will produce the predicted effects. If there is no effect in an artificial grammar learning context in adults, it does not mean that such an effect would not be present in infants or children, or in more naturalistic language learning settings, both first and second. There is some evidence that infants and children show different sensitivities and biases for learning novel phonological patterns, even in the laboratory (Gerken, Quam \& Goffman 2019). This is a question for future research.

While the present study did not show any conclusive results, they make several important contributions to phonological theory. The results pose the question of whether articulatory differences between vowels in different contexts may have any psychological benefit to learning or representations of vowel harmony patterns. If not, then it could provide support for more abstract, or 'substance-free' approaches to phonology, where phonetic underpinnings are on a different level of abstraction (Hale \& Reiss 2000). In addition, the minimal effects of phonetic

[^3]naturalness in the learnability of phonological patterns, it is clear that these are important questions to be addressed.

## References

Baković, Eric \& Colin Wilson. 2000. Transparency, strict locality, and targeted constraints. West Coast Conference on Formal Linguistics (WCCFL) 19. 43-56.
Bates, Douglas, Martin Mächler, Ben Bolker \& Steve Walker. 2015. Fitting linear mixed-effects models using \{lme4\}. Journal of Statistical Software 67. 1-48. https://doi.org/10.18637/jss.v067.i01.
Benus, Stefan \& Adamantios I. Gafos. 2007. Articulatory characteristics of Hungarian 'transparent' vowels. Journal of Phonetics 35(3). 271-300 https://doi.org/10.1016/j.woen.2006.11.002.
Blaho, Sylvia \& Dániel Szeredi. 2013. Hungarian neutral vowels: A microcomparison. Nordlyd 40. 20-40.

Boersma, Paul \& David Weenink. 2017. Praat: Doing phonetics by computer.
Carpenter, Angela. 2010. A naturalness bias in learning stress. Phonology 27(3). 345-392. https://doi.org/10.1017/S0952675710000199.
Clements, George N. \& Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In The structure of phonological representations, 213-255. 2nd edn.
Cohen, Jonathan D., Brian MacWhinney, Matthew Flatt \& Jefferson Provost. 1993. PsyScope: A new graphic interactive environment for designing psychology experiments. Behavioral Research Methods, Instruments and Computers 25. 257-271. https://doi.org/10.3758/BF03204507.
Do, Youngah. 2018. Paradigm uniformity bias in the learning of Korean verbal inflections. Phonology 35(4). 547-575. https://doi.org/10.1017/S0952675718000209.
Ebersole, Charles R., Olivia E. Atherton, Aimee L. Belanger, Hayley M. Skulborstad, Jill M. Allen, Jonathan B. Banks, Erica Baranski, et al. 2016. Many Labs 3: Evaluating participant pool quality across the academic semester via replication. Journal of Experimental Social Psychology 67. 68-82. https://doi.org/10.1016/j.jesp.2015.10.012.
FindingFive Team. 2019. FindingFive: A web platform for creating, running, and managing your studies in one place. NJ, USA: FindingFive Corporation (nonprofit). https://www.findingfive.com.
Finley, Sara. 2008. Formal and cognitive restrictions on vowel harmony. Baltimore, MD: Johns Hopkins University dissertation.
Finley, Sara. 2012. Typological asymmetries in round vowel harmony: Support from artificial grammar learning. Language and Cognitive Processes 27(10). 1550-1562. https://doi.org/10.1080/01690965.2012.660168.
Finley, Sara. 2015. Learning nonadjacent dependencies in phonology: Transparent vowels in vowel harmony. Language 91(1). 48-72. https://doi.org/10.1353/lan.2015.0010.
Finley, Sara. 2017. Learning metathesis: Evidence for syllable structure constraints. Journal of Memory and Language 92. https://doi.org/10.1016/j.jml.2016.06.005.
Finley, Sara \& William Badecker. 2009. Artificial language learning and feature-based generalization. Journal of Memory and Language 61(3). 423-437. https://doi.org/10.1016/j.jml.2009.05.002.
Finley, Sara \& William Badecker. 2012. Learning biases for vowel height harmony. Journal of Cognitive Science 13. 287-327.

Gerken, Lou Ann, Carolyn Quam \& Lisa Goffman. 2019. Adults Fail to Learn a Type of Linguistic Pattern that is Readily Learned by Infants. Language Learning and Development 15(4). 279-294. https://doi.org/10.1080/15475441.2019.1617149.
Gick, Bryan, Douglas Pulleyblank, Fiona Campbell \& Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. Phonology 23(01). 1. https://doi.org/10.1017/S0952675706000741.
Glewwe, Elenor. 2019. Bias in phonotactic learning: Experimental studies of phonotactic implicationals. Los Angeles, CA: UCLA dissertation.
Goldrick, Matthew. 2000. Turbid output representations and the unity of opacity. Northeastern Linguistics Society (NELS) 30. 231-246.
Grimm, Lisa R., Arthur B. Markman \& W. Todd Maddox. 2012. End-of-semester syndrome: How situational regulatory fit effects test performance over an academic semester. Basic and Applied Social Psychology 34(4). 376-385. https://doi.org/10.1080/01973533.2012.693427.
Hale, Mark \& Charles Reiss. 2000. "Substance abuse" and "dysfunctionalism": Current trends in phonology. Linguistic Inquiry 31(1). 157-169. https://doi.org/10.1162/002438900554334.
Kaun, Abigail R. 2004. The typology of rounding harmony. In Bruce Hayes, Robert Kirchner \& Donca Steriade (eds.), Phonetically based phonology, 87-116. Cambridge: Cambridge University Press.
Kimper, Wendell A. 2011. Competing triggers: Transparency and opacity in vowel harmony. Amherst, MA: University of Massachusetts dissertation.
Kimper, Wendell A. 2016. Asymmetrical generalisation of harmony triggers. Annual Meeting on Phonology (AMP) 3. 1-12. https://doi.org/10.3765/amp.v3i0.3662.
Lysvik, Julian Kirkeby. 2020. Where does naturalness in phonology come from? Insights from artificial language learning. Oslo, Norway: University of Oslo dissertation.
Martin, Alexander \& James Clifford White. 2021. Vowel harmony and disharmony are not equivalent in learning. Linguistic Inquiry 52(1). 227-239. https://doi.org/10.1162/ling_a_00375.
McMullin, Kevin \& Gunnar Ólafur Hansson. 2019. Inductive learning of locality relations in segmental phonology. Laboratory Phonology 10(1). 14. https://doi.org/10.5334/labphon. 150.
Moreton, Elliott \& Joe Pater. 2012a. Structure and substance in artificial-phonology learning, part I: Structure. Language and Linguistics Compass 6(11). 686-701. https://doi.org/10.1002/lnc3.363.
Moreton, Elliott \& Joe Pater. 2012b. Structure and substance in artificial-phonology learning, Part II: Substance. Language and Linguistics Compass 6(11). 702-718. https://doi.org/10.1002/lnc3.366.
Moreton, Elliott, Joe Pater \& Katya Pertsova. 2017. Phonological concept learning. Cognitive Science 41(1). 4-69. https://doi.org/10.1111/cogs. 12319.
Moreton, Elliott \& Katya Pertsova. 2017. Implicit and explicit processes in phonotactic learning. Boston University Conference on Language Development (BUCLD) 40. 277-290.
Nicholls, Michael E.R., Kellie M. Loveless, Nicole A. Thomas, Tobias Loetscher \& Owen Churches. 2015. Some participants may be better than others: Sustained attention and motivation are higher early in semester. Quarterly Journal of Experimental Psychology 68(1). 10-18. https://doi.org/10.1080/17470218.2014.925481.

Prince, Alan \& Paul Smolensky. 2004. Optimality Theory: Constraint interaction in generative grammar. Cambridge: Blackwell.
Pycha, Anne, Pawel Nowak, Eurie Shin \& Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. West Coast Conference of Formal Linguistics (WCCFL) 22. 101-113.
R Development Core Team, R. 2018. R: A language and environment for statistical computing. (Ed.) R Development Core Team. $R$ Foundation for Statistical Computing (R Foundation for Statistical Computing). R Foundation for Statistical Computing. https://doi.org/10.1007/978-3-540-74686-7.
Rebrus, Péter \& Miklós Törkenczy. 2015. Monotonicity and the typology of front/back harmony. Theoretical Linguistics 41. 1-61. https://doi.org/10.1515/tl-2015-0001.
Ringen, Catherine O. 1988. Transparency in Hungarian vowel harmony. Phonology 5(2). 327342. https://doi.org/10.1017/S0952675700002335.

RStudio Team. 2020. RStudio: Integrated Development for R. http://www.rstudio.com/.
Skoruppa, Katrin \& Sharon Peperkamp. 2011. Adaptation to novel accents: Feature-based learning in context-sensitive phonological regularities. Cognitive Science 35(2). 348-366. https://doi.org/10.1111/j.1551-6709.2010.01152.x.
Smith, Caitlin. 2016. A gestural account of neutral segment asymmetries in harmony. Annual Meeting on Phonology (AMP) 3. 1-12. http://dx.doi.org/10.3765/amp.v3i0.3663.
Szeredi, Daniel. 2016. Exceptionality in vowel harmony. New York: NYU dissertation.
Wilson, Colin. 2003a. Analyzing unbounded spreading with constraints: Marks, targets, and derivations. Ms. UCLA.
Wilson, Colin. 2003b. Experimental investigations of phonological naturalness. West Coast Conference of Formal Linguistics (WCCFL) 22. 101-114.
Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. Cognitive Science 30. 945-982.
https://doi.org/10.1207/s15516709cog0000_89.


[^0]:    * I would like to thank my research assistants: Saara Charania, Tiarra Lewis, Cristina Flores, Stella Wang, Jorge Molinera, Liza Radford, and Shelan Porter. I would also like to thank Carrie Chiasson help in producing stimuli. In addition, I would like to thank audiences at the 2021 LSA annual meeting, as well as audiences at the 2017 NowPhon meeting, the UBC Linguistics Department, and participants at Pacific Lutheran University. I assume responsibility for any and all errors in this document. Author: Sara Finley, Pacific Lutheran University (finleysr@plu.edu).

[^1]:    ${ }^{1}$ A full set of materials, including audio files, analysis code, and data can be found at: https://osf.io/mtavw.

[^2]:    ${ }^{2}$ Some models showed singular fit warnings. When this occurred, a simpler model was run in order to remove the warning. In all cases, the simpler model showed nearly identical results.

[^3]:    ${ }^{3}$ Note that Lysvik's (2020) experimental results did not support a difference between perception and production in naturalness biases in learning.

