The Role of Negative and Positive Evidence in Adult Phonological Learning

Sara Finley
Elmhurst College

The Role of Negative and Positive Evidence in Adult Phonological Learning

Abstract

One of the great mysteries of language development is how children acquire language so efficiently while adults are never able to reach the same level of proficiency. Adding to this mystery is that child learners rarely receive negative evidence regarding the nature of the grammatical structure of their language, but adults are more likely to receive and use such evidence (in classes, corrections, etc.) (Baker, 1979). The present study tests the role of negative evidence in adult language learners, who were exposed to an artificial grammar characterized by round vowel harmony, a phonological process whereby vowels agree in the feature round. Participants were exposed to either positive evidence only (Positive Evidence Condition), or both negative and positive evidence (Positive Evidence Condition). In two experiments, participants in the Positive Evidence Condition outperformed participants in the Negative Evidence Condition, specifically for test items tat measured extension of learned items to novel items. These results suggest that negative evidence may hinder adult grammatical rule learning.

The Role of Negative and Positive Evidence in Adult Phonological Learning

Sara Finley*

1 Introduction

This paper discusses the role of negative evidence in adult phonological learning. Negative evidence is defined as any type of evidence for a grammatical structure that uses what is *ungrammatical* to help the learner determine the rules for grammaticality. In typical language development, children tend to use only positive evidence to learn language. Children learn language through positive exemplars of the target language without specifically comparing grammatical and ungrammatical structures.

One puzzle about typical first language development is that children learn language in spite of a lack of negative evidence (Baker 1979). This puzzle presents a paradox to theories of language learning, particularly because many computational models of language learning make explicit use of negative evidence. Specifically, many computational models of phonological learning make use of error correction, a form of negative evidence (Boersma and Hayes 2001, Tesar and Smolensky 1998). A question that remains is how children learn their native language, despite a lack of negative evidence or explicit error correction.

An interesting piece to this puzzle is that while children do not receive or use negative evidence when learning a first language, second language learning contains many instances of negative evidence. For example, teachers correct their students in foreign and second language classes. This correction comes in the form of exams, written assignments and verbal corrections. However, despite the large amount of negative evidence in second language learning situations, second language learning falls short to first language learning in terms of the proficiency of complex, productive grammatical structure. Adults rarely, if ever, achieve native-level competence for a second language, and this is particularly true for grammatical rules (Johnson and Newport 1989).

The question that is addressed in the present paper is whether negative evidence actually hinders language learning in adult human learners. If negative evidence hinders learning rather than helps learning, then the fact that negative evidence is unavailable to the child learner can help to explain some of the differences between children and adults in language learning. The present set of experiments tests this hypothesis using an artificial grammar learning paradigm with adult participants. Adult native English speakers were exposed to a miniature language that consisted of a single phonological pattern, vowel harmony. In vowel harmony, all of the vowels in a word share the same phonological feature. For example, in a simplified version of Turkish, if the vowel of the stem of a word is front, all suffix vowels will also be front (e.g., /ip/ 'rope' takes the front suffix /-ler/, Clements and Sezer 1982).

There are two reasons that vowel harmony was chosen as a test case for studying the role of negative evidence in adult phonological learning. First, vowel harmony is a phonological pattern that is explicitly taught in second and foreign language instruction (when the language in question contains a productive vowel harmony system). In contrast, children learning a vowel harmony language acquire vowel harmony without any explicit instruction. Second, artificial grammar learning experiments using vowel harmony have been readily conducted with successful learning, demonstrating that implicit learning of a miniature vowel harmony language is possible with positive evidence (Finley and Badecker 2008, 2009, Pycha et al. 2003).

In the present study, participants were exposed to a miniature language that demonstrated a back/round vowel harmony pattern in which a suffix alternated between a back/round vowel (/u/) and a front/unround vowel (/i/) depending on the back/round features of a stem vowel. Half of participants were provided with only positive evidence of the harmony pattern, while the other half of participants were provided with both positive and negative evidence for the vowel harmony

*The author would like to thank Elissa Newport, Paul Smolensky, Patricia Reeder, Neil Bardhan, Carrie Miller, Emily Kasman, Kelly Johnston, Lily Schieber, Anna States, and members of the Aslin-Newport lab. This research was supported in part by NIH grant DC00167 to E. Newport, HD37082 to R. Aslin & E. Newport, and NIH training grant T32DC000035.

pattern. Participants learned and generalized the harmony pattern in both the Positive Evidence Condition and the Negative Evidence Condition. However, the rate of generalization of the harmony pattern was higher in the Positive Evidence Condition than in the Negative Evidence Condition. In Experiment 2, the advantage of the Positive Evidence Condition persisted even when the number of positive exemplars in the Negative Evidence Condition was increased, and an additional gender cue in the Positive Evidence Condition was removed. These results suggest that negative evidence hinders the learning of novel phonological patterns.

2 Experiment 1

The present experiment tested the role of negative evidence in learning novel phonological patterns.

2.1 Participants

Forty adult students and affiliates of the University of Rochester participated in Experiment 1. All participants were adult native English speakers, with no knowledge of a vowel harmony language, and had not participated in a previous vowel harmony learning experiment. All participants were paid $10 for their participation.

2.2 Design and Procedure

Participants were randomly assigned to one of two conditions: a Positive Evidence Condition and a Negative Evidence Condition. In both conditions, all 'words' of the miniature language consisted of three CVCVCV syllables. The first two syllables in all words were harmonic for back and round features. The first two syllables consisted of the following harmonic vowel pairs: /i,i/, /i,e/, /e,e/, /e,i/, /u,u/, /u,o/, /o,o/ and /o,u/. Because these syllables followed the vowel harmony pattern, the first two syllables of all exposure words did not contain disharmonic vowel pairs such as */i,u/ or */o,e/.

The final syllable of the tri-syllabic items alternated between /-mi/ and /-mu/. This alternation was designed as a pseudo-suffix in order to simulate stem-controlled harmony that is commonly found in natural harmony languages (Backovic 2000). The training set in the Positive Evidence Condition contained only harmonic items and were termed 'positive evidence' items. For these items, if the vowels in the first two syllables contained front/unround vowels /i/ or /e/, /-mi/ appeared in the final syllable (e.g., /kipemi/). If the vowels in the first two syllables contained back/round vowels /u/ or /o/, /-mu/ appeared in the final syllable (e.g., /bodomu/).

The training set used in the Negative Evidence Condition contained fully harmonic 'positive evidence' items (36 items total) in addition to disharmonic, 'negative evidence' items (12 items total). The final vowel of the negative evidence items was always disharmonic. The suffix containing the unround vowel, /-mi/, appeared in words that contained round vowels (/u/ and /o/) in the first two syllables (e.g., /bodomi/). The suffix containing the round vowel, /-mu/, appeared in words that contained unround vowels (/i/ and /e/) in the first two syllables (e.g., /kipemu/). In order to distinguish between harmonic and disharmonic items in the Negative Evidence Condition, the harmonic items were recorded using a male voice, while the disharmonic items were recorded using a female voice. In order to balance the use of multiple voices in the Negative Evidence Condition, harmonic items in the Positive Evidence Condition were recorded using both male and female voices.

Participants in both conditions were exposed to 48 items (36 items in the male voice, and 12 items in the female voice) repeated 5 times each (in a random order), with the items that were recorded using a female voice randomly interspersed with items that were recorded using a male voice. In the Positive Evidence Condition, participants were told that they would be listening to words from a language that they had never heard before, and that the words would be presented in two different voices, one male and one female. In the Negative Evidence Condition, participants were told that they were listening to a language that they had never heard before, and that two speakers would help them learn the language. The male speaker says words that are in the language, while the female speaker says words that are not in the language. Participants were given a

memory tool: male = language, female = not language.

Following exposure, all participants were given a two-alternative forced-choice test that contained the same set of items for both training conditions. The test items contained 12 items that appeared in the training set (labeled 'Old' items), and 12 items that did not appear in the training set (labeled 'New' items). The use of novel items allowed a test for generalization of the harmony pattern to new items, ruling out the possibility that learners simply memorize the words in the training set, rather than learning a productive pattern. The test items contained a choice between a harmonic item and a disharmonic item (e.g., /bede-mi/ vs. */bede-mu/), and differed only in the final syllable /-mi/ vs. /-mu/. If negative evidence hinders learners from forming abstract grammatical rules such as vowel harmony, participants in the Positive Evidence Condition will perform more accurately on New Items than participants in the Negative Evidence Condition.

Participants were told that they would hear two words, one from the language that they had just heard, and one not from the language, and that their job was to select the word that was most likely to belong to the language that they had just heard. All test items were presented using the male voice. Responses were coded using a key press ('a' for the first item, and 'l' for the second item). The entire experiment, presented in PsyScope X (Cohen et al. 1993), took approximately 15 minutes.

2.3 Materials

Two adult native English speakers recorded all the stimuli in a sound-attenuated booth. The speakers were not aware of the design of the experiment. Both speakers were told to produce the materials in a clear voice, pronouncing all vowels. This was done in order to reduce the amount of vowel reduction. Stress for all items was placed on the first syllable of the tri-syllabic item. All items were normalized for loudness.

All stimuli were of the form CVCVCV with the final syllable having alternating between the suffixes /-mi/ and /-mu/. All other segments in the stimuli were drawn from a set of consonants /p, t, k, b, d, g, m, n/ and vowels /i, e, u, o/. The first two syllables (stems) were of the form CVCV and were created semi-randomly, with the caveat that any word closely resembling a known English word was not used in the exposure or test sets. Use of each vowel was kept consistent over different vowel pairs in order to minimize the number of items having multiple identical vowels in a row (though this did occur in some of the items). Examples of exposure and test stimuli can be found in Table 1.

Condition	Exposure	Old Items	New Items
Positive Evidence	**Male Voice** podomu degimi gibemi	**Male Voice:** negemi vs. *negemu *tonomi vs. tonomu bimimi vs. *bimimu *podomi vs. podomu gibemi vs.gibemu *punumi vs. punumu	**Male Voice:** kenemi vs. *kenemu *modumi vs. modumu tidimi vs. *tidimu *gutomi vs. gutomu kipemi vs.kipemu *mogomi vs. mogomu
	Female Voice podomu gibemi		
Negative Evidence	**Male Voice** podomu degimi gibemi	**Male Voice:** negemi vs. *negemu *tonomi vs. tonomu bimimi vs. *bimimu *podomi vs. podomu gibemi vs.gibemu *punumi vs. punumu	**Male Voice:** kenemi vs. *kenemu *modumi vs. modumu tidimi vs. *tidimu *gutomi vs. gutomu kipemi vs.kipemu *mogomi vs. mogomu
	Female Voice podomi gibemu		

Table 1: Sample exposure and test items.

2.4 Results

The proportion of harmonic responses was recorded for all participants, and is shown in Figure 1. We compared both the Old and the New Items to chance (50%) via Bonferroni corrected one-

sample *t*-tests in order to test whether participants learned the harmony pattern. Participants in the Negative Evidence Condition demonstrated above-chance performance for both Old Items, with a mean of 0.62, *CI±* 0.088, $t(19)= 3.01$, $p < 0.05$, as well as New Items, with a mean of 0.61, *CI±* 0.068, $t(19)= 3.56$, $p < 0.01$. This suggests that participants learned the harmony pattern in the Negative Evidence Condition. Participants in the Positive Evidence Condition demonstrated above-chance performance for both Old Items, with a mean of 0.70, *CI±* 0.083, $t(19)= 5.02$, $p < 0.001$, as well as New Items, with a mean of 0.71, *CI±* 0.070, $t(19)= 6.30$, $p < 0.001$. This suggests that participants learned the harmony pattern in the Positive Evidence Condition.

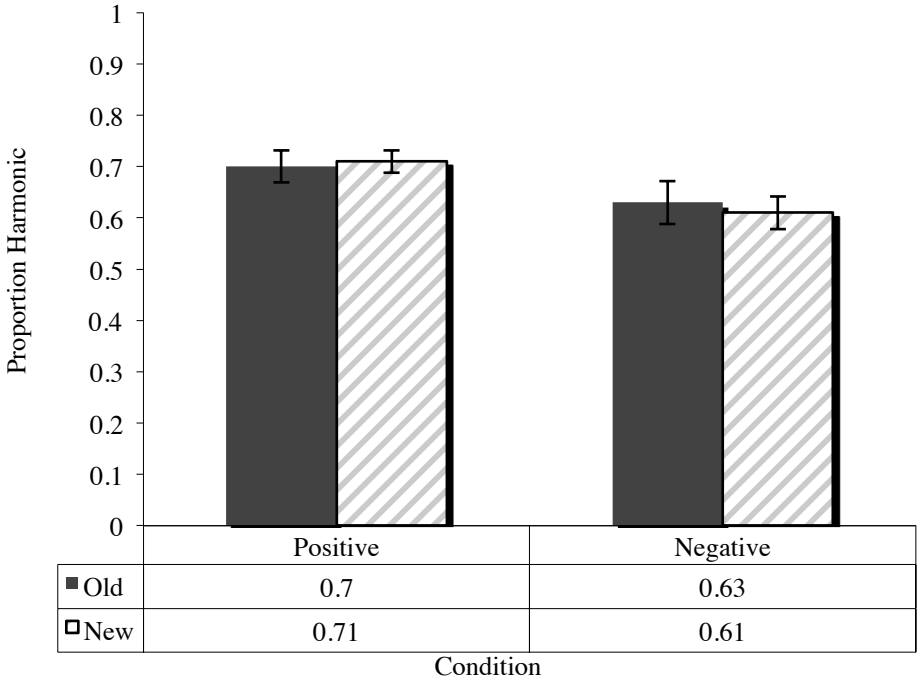

Figure 1: Experiment 1 Results: Means and Standard Errors.

The crucial test of the hypothesis in the present experiment is whether participants are more likely to generalize the harmony when they have negative evidence for the harmony pattern. To answer this question, we compared responses to New Items between the Positive Evidence Condition and the Negative Evidence Condition. We found that participants in the Positive Evidence Condition showed greater generalization of the harmony pattern (mean = 0.71 vs. 0.61, *CI±* 0.094, $t(38)= 2.08$, $p= 0.044$), suggesting that participants in the Positive Evidence Condition showed greater generalization of the harmony pattern than participants in the Negative Evidence Condition.[1]

2.5 Discussion

Participants in Experiment 1 learned the harmony pattern, both in the Negative Evidence Condition and in the Positive Evidence Condition. This suggests that it is possible to learn a simple phonological pattern with both negative and positive evidence. However, participants in the Positive Evidence Condition showed higher levels of performance for New Items compared to participants in the Negative Evidence Condition. This suggests that negative evidence may hinder the learning process, specifically for forming general, rule-like patterns.

It is unclear from Experiment 1 whether participants in the Negative Evidence Condition were

[1]There was a marginal overall effect between the Positive Evidence Condition and the Negative Evidence Conditions via a 2x2 ANOVA, $F(1, 38) = 3.44$, $p = 0.071$.

at a disadvantage compared to participants in the Positive Evidence Condition simply because there were fewer positive evidence tokens in the Negative Evidence Condition. Participants in both conditions of Experiment 1 heard 48 words repeated 5 times each. This means that the participants in the Positive Evidence Condition heard 48 harmonic tokens, while the participants in the Negative Evidence Condition heard only 36 harmonic tokens.

In addition, participants in the Positive Evidence Condition heard positive evidence in both a male and a female voice. There is evidence that infants are more likely to learn a phonological pattern when that pattern is presented in multiple voices (Richtsmeier 2010). While this pattern did not extend to adults (Richtsmeier 2011), it is worth controlling for in a follow-up experiment.

3 Experiment 2

Experiment 2 further extended the results of Experiment 1 by testing the hypothesis that negative evidence hinders learning, even when there are equal numbers of positive evidence items in both conditions, and when all positive evidence items are presented in the male voice. Experiment 2 increased the number of positive evidence items in the Negative Evidence Condition and removed the items recorded in a female voice in the Positive Evidence Condition.

3.1 Participants

Forty adult students and affiliates of the University of Rochester participated in Experiment 2. All participants were adult native English speakers, with no knowledge of a vowel harmony language, and had not participated in a previous vowel harmony learning experiment (including Experiment 1). All participants were paid $10 for their participation.

3.2 Design and Procedure

Experiment 2 used essentially the same design as Experiment 1 with the following differences. First, the Positive Evidence Condition heard only male voice items during the exposure phase. Second, the exposure phase of the Negative Evidence Condition had the same number of positive, harmonic items as the Positive Evidence Condition (48 positive evidence items in each condition). If negative evidence hinders learning, participants in the Negative Evidence Condition should perform worse than participants in the Positive Evidence Condition, despite having the same number of positive exemplars (in addition to the negative exemplars).

3.3 Materials

Experiment 2 used identical materials as Experiment 1, except that there were no items using the female voice in the Positive Evidence Condition.

3.4 Results

The proportion of harmonic responses was recorded for all participants, and is provided in Figure 2. To test whether participants learned the harmony pattern, we compared both the Old and the New Items to chance (50%) via Bonferroni corrected one-sample *t*-tests. Participants in the Negative Evidence Condition showed above-chance performance for Old Items, with a mean of 0.61, $CI\pm$ 0.078, $t(19)= 2.90$, $p < 0.05$, but only marginally significant performance for New Items after Bonferroni correction, with a mean of 0.59, $CI\pm$ 0.081, $t(19)= 2.28$, $p = 0.07$. This suggests that participants learned the harmony pattern in the Negative Evidence Condition. Participants in the Positive Evidence Condition showed above-chance performance for both Old Items, with a mean of 0.69, $CI\pm$ 0.079, $t(19)= 5.07$, $p < 0.001$ as well as New Items, with a mean of 0.70, $CI\pm$ 0.060, $t(19)= 6.86$, $p < 0.001$. This suggests that participants learned the harmony pattern in the Positive Evidence Condition.[2]

[2]There was a significant overall effect between the Positive Evidence Condition and the Negative Evidence Conditions via a 2x2 ANOVA, $F(1, 38) = 4.84$, $p = 0.034$.

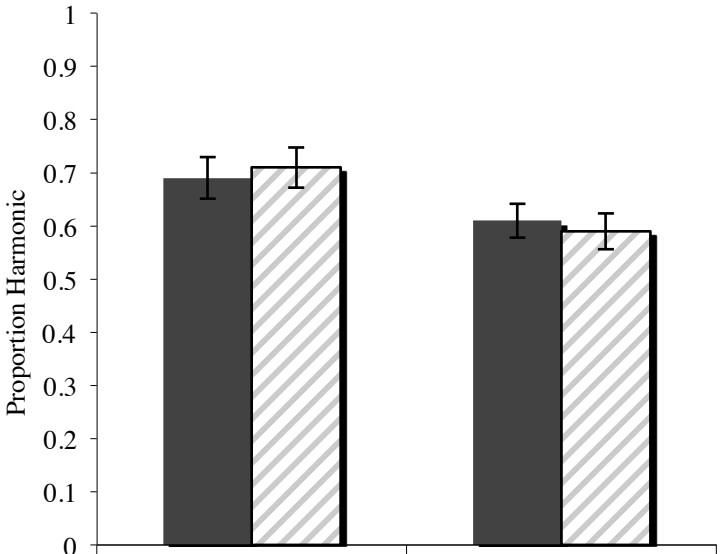

Figure 2: Experiment 2 Results: Means and Standard Errors.

The crucial test of the hypothesis in the present experiment is whether participants generalize the harmony pattern better when the learners have negative evidence for the harmony pattern. To answer this question, we compared responses to New Items between the Positive Evidence Condition and the Negative Evidence Condition. We found that participants in the Positive Evidence Condition showed greater generalization of the harmony pattern (mean = 0.71 vs. 0.59, $CI\pm$ 0.094, $t(38)= 2.08$, $p= 0.034$), suggesting that participants in the Positive Evidence Condition showed greater generalization of the harmony pattern than participants in the Negative Evidence Condition.

3.5 Discussion

Participants in the Positive Evidence Condition continued to demonstrate an increased extension of the harmony pattern, even when the Positive Evidence Condition only heard positive evidence in one voice, and when participants in the Negative Evidence Condition heard as many positive exemplars as participants in the Positive Evidence Condition. This supports the hypothesis that negative evidence impedes grammar learning.

4 General Discussion and Conclusions

The present study presented two artificial grammar learning experiments using adult native English speakers. Participants were exposed to a vowel harmony pattern. Following exposure, the participants were given a two-alternative forced choice task between harmonic and disharmonic items. These items included items that were heard in the exposure set (Old) as well as novel items that had never been heard before (New). Half of the participants heard only positive exemplars, tokens that fit the vowel harmony pattern, while the other half of participants heard both positive and negative exemplars (negative evidence). Participants in the Positive Evidence Condition outperformed the participants in the Negative Evidence Condition on the New Items. This suggests that negative evidence impedes the learning of novel phonological patterns.

This result is in line with the fact that adults are more likely to receive negative evidence in learning a second language than children are in learning their first language. However, despite

receiving negative evidence, children outperform adults in language learning. Specifically, children outperform adults on learning abstract grammatical patterns (Johnson and Newport 1989). This is also in line with the present results because the major differences between the Negative Evidence Condition and the Positive Evidence condition were found in the New Items. New Items were the items that demonstrated learning a grammatical pattern (as opposed to memorizing the exposure items).

The present results support a view of first language learning in which only positive evidence is needed, and that the learner must filter out ungrammatical items, even when they are explicitly given as negative evidence (Perfors, Tenenbaum, and Wonnacott 2010). It may be that children have a greater ability to filter out negative evidence than adults, and it is this ability that helps children outperform adults on learning the grammar of a language.

The present study opens up several avenues for future research. First, the present experiment explored adult learning. If children are better at filtering out negative evidence in language learning than adults, there may not be any hindrance to grammar learning when children are given both negative and positive evidence. Second, more research is needed to discover the mechanism for which negative evidence might hinder grammar learning, and how learners cope with this hindrance. For example, the participants in the Negative Evidence Condition heard negative evidence through a gender cue (the female voice represented the negative evidence), and these negative evidence items were randomly interspersed throughout the exposure phase. It is possible that negative evidence may not have as much of a hindering effect when negative evidence is presented in a more structured manner. In the case of negative evidence in real-world second language learning situations, negative evidence is often interspersed with positive evidence, suggesting that the random presentation did, in part, model real-life experiences with negative evidence. However, the present study was not designed to simulate computational models of language learning. Computational models that make use of negative typically receive error correction and supervision. In the present experiment, there was no error correction; learning was entirely passive. Future research will work to demonstrate whether human learning proceeds in a similar fashion to computational models of learning.

The present study demonstrated that negative evidence might have a hindering effect in learning novel grammars. This finding may help to explain Baker's paradox that children, despite only receiving positive evidence (and ignoring negative evidence), become proficient language learners, while adults, who receive both negative and positive evidence, fail to achieve native-like language proficiency. The present study may also help to design curriculum for adult language learners. If negative evidence makes it more difficult to form abstract grammatical rules, second language teachers may work to reduce the amount of negative evidence that second language learners receive in the classroom.

References

Bakovic, Eric. 2000. Harmony, dominance, and control. Doctoral dissertation, Rutgers University.

Baker, Carl Lee. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10:533–581.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:45–86.

Clements, George, and Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In *The Structure of Phonological Representations*, eds. H. van der Hulst and N. Smith, 213–255. Dordrecht: Foris.

Cohen, Jonathan D., Brian MacWhinney, Matthew Flatt, and Jefferson Provost. 1993. PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers* 25:257–271.

Finley, Sara, and William Badecker. 2008. Substantive biases for vowel harmony languages. In *Proceedings of the West Coast Conference of Formal Linguistics 30*, eds. N. Abner and J. Bishop, 168–176.

Finley, Sara, and William Badecker. 2009. Artificial grammar learning and feature-based generalization. *Journal of Memory and Language* 61:423–437.

Johnson, Jacqueline, and Elissa Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21:60–99.

Perfors, Amy, Josh Tenenbaum, and Elizabeth Wonnacott. 2010. Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language* 37:607–642.
Pycha, Anne, Pawel Nowak, Eurie Shin, and Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In *Proceedings of the West Coast Conference on Formal Linguistics 22*, eds. G. Garding and M. Tsujimura.
Richtsmeier, Peter. 2011. Adults learn phonotactic probabilities from word-types not word-tokens. *Laboratory Phonology* 2:1–33.
Richtsmeier, Peter, LouAnn Gerken, Lisa Hoffman, and Tiffany Hogan. 2010. Statistical frequency in perception affects children's lexical production. *Cognition* 111:372–377
Tesar, Bruce, and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29:229–268.

Department of Psychology
Elmhurst College
Elmhurst, IL 60126
finleys@elmhurst.edu