Exploring the relationship between learnability and linguistic universals

Anna N. Rafferty (rafferty@cs.berkeley.edu)
Computer Science Division, University of California, Berkeley, CA 94720 USA

Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California, Berkeley, CA 94720 USA

Marc Ettlinger (marc@northwestern.edu)
Department of Communication Sciences and Disorders
Northwestern University, Evanston, IL 60208 USA

Abstract

Greater learnability has been offered as an explanation as to why certain properties appear in human languages more frequently than others. Languages with greater learnability are more likely to be accurately transmitted from one generation of learners to the next. We explore whether such a learnability bias is sufficient to result in a property becoming prevalent across languages by formalizing language transmission using a linear model. We then examine the outcome of repeated transmission of languages using a mathematical analysis, a computer simulation, and an experiment with human participants, and show several ways in which greater learnability may not result in a property becoming prevalent. Both the ways in which transmission failures occur and the relative number of languages with and without a property can affect whether the relationship between learnability and prevalence holds. Our results show that simply finding a learnability bias is not sufficient to explain why a particular property is a linguistic universal, or even frequent among human languages.

1 Introduction

A comparison of languages around the world reveals that certain properties are far more frequent than others, which are taken to reflect linguistic universals (Greenberg, 1963; Comrie, 1981; Croft, 2002). Understanding the origins of linguistic universals is an important project for linguistics, and understanding how they relate to human cognitive processes is an important project for cognitive science. One prominent explanation for the existence of these patterns is the presence of cognitive biases that make certain properties of language more easily learned than others (Slobin, 1973; Wilson, 2003; Finley & Badecker, 2007; Wilson, 2006). Under this hypothesis, certain properties are common across languages because they are more easily learned than others (a *learnability bias*) and are therefore more likely to be maintained when a language is passed from one generation to the next. These universals generally reflect tendencies, rather than properties that are present in each and every language (Croft, 2002).

Recent work in psycholinguistics has provided support for a relationship between learnability biases and the properties that are prevalent in human languages. A number of studies have shown that certain common phonological patterns, such as vowel harmony, voicing agreement and final devoicing are, indeed, more learnable than other unattested patterns (Finley & Badecker, 2007; Moreton, 2008; Becker, Ketrez, & Nevins, 2011). Based on these findings, it is tempting to argue that learnability biases alone might account for the prevalence of these properties in human languages. However, this argument assumes that more accurate learning of a language with a certain property is sufficient for that property to become widespread across languages and does not account for why a property might be prevalent but not universal across languages.

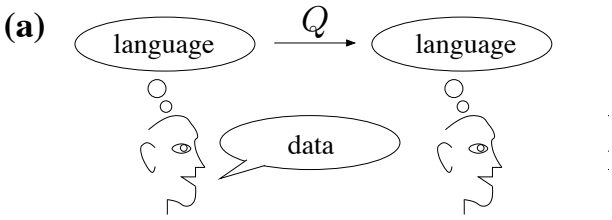
In this paper, we examine the assumption that greater learnability is sufficient for a property to become prevalent. We formalize language transmission using a simple linear model, and then show two basic scenarios in which greater learnability for a particular language does not result in that language becoming prevalent. We first perform a mathematical analysis to show that one way this can occur is for errors in transmission to favor particular lan-

guages over others. We next use a simulation to show another scenario in which greater learnability can fail to result in a dominant pattern: when the number of alternative languages is large. We conduct two experiments with human participants to illustrate the occurrence of this second scenario in the case of a particular property of human language, vowel harmony.

2 Linking Learnability and Transmission

Languages change over time due to transmission from generation to generation (e.g., Labov, 2001). Our goal is to understand how long-term trends of language change are related to cognitive, perceptual, and production biases observed in a single instance of transmission. We begin by formalizing transmission using a general mathematical model in order to uncover what long term trends emerge given that certain languages are more likely to be accurately transmitted than others.

We use a linear model of cultural transmission, in which it is assumed that each person learns a language from utterances produced by one person in the previous generation. This linear model of transmission has many specific instantiations in the literature on language evolution, such as the iterated learning model (Kirby, 2001; Griffiths & Kalish, 2007) or the replicator dynamics (Schuster & Sigmund, 1983; Komarova & Nowak, 2003). To specify this model, we first define the set of possible languages, denoted H. Each element $h \in H$ is one possible language. Transmission occurs when a new member of the population receives linguistic data (a set of utterances) from another member of the population and learns a language $h \in H$. We assume transmission occurs only from one person to another person, and that each person learns only one language. For example, someone who knows language j might speak to another member of the population, and based on hearing those utterances, the learner might also learn the language j. Alternatively, the learner might learn another language: The learner might not have heard enough language to fully specify j as the language or might have misheard something, and thus simply infers another language i that is consistent with the data she or he heard. More generally, we assume that for all $i, j \in H$, q_{ij} is the probability that someone will learn language i from someone who knows language j. These can be encoded in a transition matrix Q where the (i, j)th entry of the matrix corresponds to q_{ij} (see Figure 1).



(b)
$$Q = \begin{bmatrix} 0.90 & 0.05 & 0.07 \\ 0.04 & 0.76 & 0.08 \\ 0.06 & 0.19 & 0.85 \end{bmatrix}$$

(c)
$$\pi = \begin{bmatrix} 0.39 \\ 0.20 \\ 0.41 \end{bmatrix}$$

Figure 1: (a) A general model of the cultural transmission of languages. A language is passed from one learner to another, and the matrix Q encodes the probability a learner will learn a particular language i from someone who knows language j. (b) An example transition matrix Q with three states. (c) The solution to the eigenvector equation $Q\pi = \pi$ for this transition matrix. π gives the equilibrium probability that a learner will learn a particular language when languages are transmitted via a process that has transition matrix Q.

Using this framework, we can formally define learnability biases and determine whether a learnability bias for some property necessarily implies that this property will be present in the majority of languages. As mentioned previously, we define learnability bias to mean that one type of language is more likely to be transmitted accurately to the next generation than another; this is similar to the notion of "cognitive bias" discussed in Wilson (2003) and is what is tested in experiments. Formally, a learnability bias for some language i over some other language j means that $q_{ii} > q_{jj}$. For example, one might expose one group of learners to language i and another group to language j. If more learners in the first group accurately learned the language they were exposed to, this would indicate a learnability bias for language i over language j.

We can extend the idea of a learnability bias to a property of a language, rather than a specific language, by applying a similar definition to sets of languages. Imagine there are two sets of languages, H_1 and H_2. These sets might be defined by classifying

all languages with a particular property in H_1 and all languages without the property in H_2. One way of defining a learnability bias that favors a particular property is for each language with that property to be more likely to be transmitted successfully than each language without that property. That is, for all possible pairs $i \in H_1$ and $j \in H_2$, $q_{ii} > q_{jj}$. This would indicate a general learnability bias for languages in H_1 over languages in H_2.

Using this definition of a learnability bias, we can determine whether such a bias is sufficient to establish that the property will be present in the majority of languages. That is, if H_1 denotes the languages with the property of interest, we want to determine whether a learnability bias for languages in H_1 implies that after many generations, the majority of the languages in the population will be in H_1 and not in H_2. We can determine the consequences of many instances of language transmission in this model by appealing to existing results on the equilibrium of this linear dynamical system. As mentioned above, this linear transmission model is related to two kinds of models that have been used to study language evolution: If we assume that learners are organized in a chain, this linear model is called iterated learning (Kirby, 2001); alternatively, if we assume that there are an infinite number of learners in the population, the model is called the replicator dynamics (Schuster & Sigmund, 1983). In either case, the probability that a learner will learn language h, assuming the population has reached equilibrium, is given by the solution to the eigenvector equation $Q\pi = \pi$, normalized such that $\sum_{i=1}^{n} \pi_i = 1$ (for details, see Griffiths & Kalish, 2007). For languages in H_1 to occur the majority of the time, it thus must be the case that $\sum_{h \in H_1} \pi_h > \sum_{h \in H_2} \pi_h$.

We can now identify one context in which a learnability bias is not sufficient to ensure that a property will appear in the majority of languages. Consider the example transition matrix Q shown in Figure 1 (b). Let $H_1 = \{s_1\}$ and $H_2 = \{s_2, s_3\}$, where each state s_i represents a distinct language. We have that $q_{11} > q_{ii}$ for all $i \in H_2$: each state in H_2 has a lower self transition probability than state s_1, the only state in H_1. Thus, we have a learnability bias for state s_1 over all states in H_2. However, the eigenvector π shown in Figure 1 (c) indicates that the equilibrium of this system, which will be reached after languages are transmitted from person to person many times, favors state s_3 over the other states. Overall, $\sum_{h \in H_1} \pi_h = 0.39$ while $\sum_{h \in H_2} \pi_h = 0.61$: most of the learners will learn a language in H_2.[1]

Intuitively, this result comes from the fact that transmission failures tend to favor languages in H_2. A learner who learns from someone who speaks a language i in H_2 will rarely learn the language in H_1, although she may learn a different language than i in H_2. This pattern of transmission failures overwhelms the learnability bias that the language in H_1 has over the languages in H_2. Note that this pattern holds even given that $q_{1i} > q_{i1}$ for all $i \in H_2$, another common criterion for a learnability bias.

This result implies that if the linear transmission model is an accurate model for understanding human language evolution, then it is not sufficient to compare how accurately languages are maintained over a single generation in order to predict what trends will emerge after many generations. Instead, one must also look at what happens when languages are not maintained accurately. The ways in which mutations occur may be as important as the relative fidelities of transmission in determining long term trends. When one only looks for a learnability bias, the rate of different mutations is not accounted for, leaving open the possibility that predictions about long term trends will be incorrect.

3 Simulating Language Transmission

In the previous section, we used a simple linear transmission model to identify one context in which a learnability bias is not sufficient for languages with a certain property to become prevalent. We now explore a second context in which a learnability bias is not sufficient to guarantee that languages with a particular property become prevalent, using a simulation of language transmission. We use an iterated learning model in which our representation of language is inspired by the principles and parameters approach (Chomsky & Lasnik, 1993). Rafferty, Griffiths, and Klein (2009) present a model similar to the one we consider here and show that compa-

[1] While one might try to resolve this issue by collapsing all languages in H_2 into a single state in the Markov chain, such a transformation is possible only in cases where $q_{ij} = q_{ik}$ for all languages $j,k \in H_2$ and $i \notin H_2$ (Burke & Rosenblatt, 1958; Kemeny & Snell, 1960).

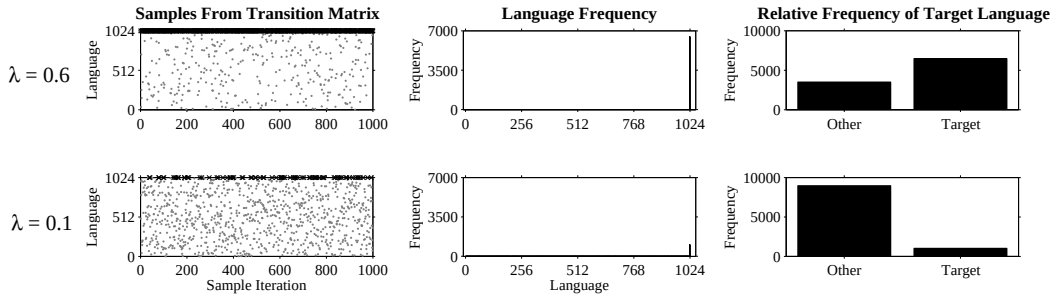


Figure 2: Model results for the frequency of the target language based on adjusting the bias towards that hypothesis. The rows in the above figure correspond to two possible values of λ; larger λ results in a higher prior probability on the target language. The leftmost column shows $1,000$ samples from the transition matrix, with black x marks corresponding to occurrences of the target language. The middle column corresponds to the frequency of each language in the full $10,000$ samples; the rightmost bar in each figure corresponds to the target language. The rightmost column shows the frequency of the target language versus all other languages for the same $10,000$ samples.

rable results hold using other representations of language, such as those based on optimality theory.

In order to define the transition matrix Q, we need to specify the process by which learners select a language. We assume that learners are Bayesian, meaning that they infer a language h based on the data d that they receive according to Bayes' rule. The *posterior probability* assigned to h after observing d is $p(h|d) \propto p(d|h)p(h)$, where $p(d|h)$ (the *likelihood*) indicates the probability of d being generated from h, and $p(h)$ (the *prior*) indicates the extent to which the learner was biased towards h before observing d. If we assume learners select hypotheses with probability equal to their posterior probability, we obtain a transition matrix Q with entries

$$\begin{aligned} q_{ij} &= p(h^{(t+1)} = i | h^{(t)} = j) \\ &= \sum_d p(h^{(t+1)} = i|d) p(d|h^{(t)} = j) \end{aligned}$$

where $h^{(t)}$ and $h^{(t+1)}$ are the languages of learners at iterations t and $t+1$ respectively.

To represent languages, we use binary vectors of length N. Each place corresponds to the setting for a particular parameter. We consider one particular setting of the parameters to be the target language and include a learnability bias for this language in the model; we then look at whether this language is more prevalent than other languages after many transmissions. In the iterated learning model that we use, learners are organized into a chain, with each learner learning from data generated by the previous learner (Kirby, 2001). The previous learner generates k pieces of data that match her or his language. These pieces of data each specify the correct parameter setting for one of the properties represented by the binary vector. The other $N-k$ properties are left unspecified in the data given to the next learner.

In order to define the transition probability between languages, we need to define the two terms in Bayes rule: the prior $p(h)$ and the likelihood $p(d|h)$. Intuitively, the prior probability distribution over languages corresponds to how much evidence is required for the learner to learn each hypothesis. If one hypothesis has a very high prior probability, only a small amount of evidence will be required to convince the learner that that hypothesis is the correct one. By controlling the prior probability of the target language versus the other languages, we can manipulate the learnability bias for the target language. We thus set the prior probability of the target language to λ and then divide the remaining probability mass of $1-\lambda$ uniformly across all of the languages (including the target language). The parameter λ thus controls the strength of the learnability bias for the target language, but this language is always favored for any λ greater than 0.

The likelihood $p(d|h)$ reflects the probability that a given hypothesis h would produce data d. We assume d is a string of length N that contains 0s, 1s, and ?s. A '?' in the ith position means that no information was given about the ith property. We also assume there is a probability ε that the chosen language will not match the data at each position; that is, with probability ε, the language chosen by the

learner will have a 1 in the ith spot if the data had a 0 in that spot. This gives:

$$p(d|h) = \prod_{i=1,d_i\neq ?}^{N} \varepsilon^{I(h\vdash d_i)}(1-\varepsilon)^{I(h\nvdash d_i)}$$

where $h \vdash d_i$ means that h has the same setting of the ith property as d_i.

Given these specifications for the prior and the likelihood, we can calculate the $2^N \times 2^N$ transition matrix and sample from this matrix to simulate a sequence of learners each learning a language from the utterances produced by the previous learner. We let $N = 10$ and $k = 5$. As shown in Griffiths and Kalish (2007), in this model – iterated learning with Bayesian learners – the equilibrium π is simply the prior distribution $p(h)$. The distribution over languages is thus unaffected by the error parameter ε; this parameter only affects the time to reach equilibrium (Rafferty et al., 2009). We present results using $\varepsilon = 0.25$. Figure 2 shows how relative frequency of the target language is affected by changing the parameter λ, using $\lambda = 0.6$ and $\lambda = 0.1$. Frequencies are based on taking $11,000$ samples from the matrix and discarding the first $1,000$ to ensure that the population had reached equilibrium.

The middle column of Figure 2 shows that the frequency with which learners chose the target language was greater than that of the other languages for both values of λ. This is consistent with the target language having a higher prior probability than other languages. However, depending on the strength of the bias, this language may still not be chosen the majority of the time, as shown in the rightmost column of Figure 2. When λ is large, its probability overwhelms that of its competitors. However, if λ is relatively small, the combined frequencies of all other languages exceed that of the target language. Thus, despite being favored by a learnability bias, the target language is not chosen by the majority of learners. Like the previous example, this simulation demonstrates that learnability biases may not always lead to accurate prediction of long term trends. More specifically, it highlights that one must consider the size of the comparison set: If there are many alternate possible languages, learners may tend to learn one of these languages even if some particular language with a learnability bias is more frequent than any other given individual language.

4 Language Transmission in the Lab

While we have shown two scenarios in which a simple linear transmission model does not predict that learnability biases will necessarily lead to linguistic universals, human learners are not necessarily consistent with this model and could follow a different pattern. Thus, we conducted two experiments to determine if the same dissociation between individual bias and long-term change can be shown when teaching human learners an artificial grammar. In Experiment 1, we establish a learnability bias for a linguistic pattern that is common in the world's languages over an arbitrary pattern. In Experiment 2, we explore what happens when a language with the common pattern is transmitted multiple times among learners in the lab. Each learner learns a language and then produces data from this language to teach the next learner. By examining the languages that emerge after several transmissions, we will show that the learnability bias in Experiment 1 does not translate to the pattern becoming widespread across the learned languages in Experiment 2. This pattern is an instance of the scenario in which the many alternative languages overwhelm the language with the learnability bias.

In our experiments, we use the property of vowel harmony. Relatively common across the world's languages (van der Hulst & van de Weijer, 1995), vowel harmony is a linguistic pattern wherein the vowels in words in a language must share some phonological feature. For example, in Turkish, the plural suffix is *-lar* in *bash-lar* 'heads', but *-ler* in *bebek-ler* 'babies' so as to adhere to the requirement that words are front-back harmonic. In the former, both vowels are back vowels and in the latter, both vowels are front vowels. Harmony is well-suited for use in this case because English speakers have no familiarity with vowel harmony from their native language input and because previous work has shown that typologically attested vowel harmony patterns are generally more easily learned (Moreton, 2008; Finley & Badecker, 2009).

5 Experiment 1: Establishing a Bias

5.1 Methods

Participants. There were 40 participants who received either monetary compensation or course

credit for their participation. All were native speakers of English.

Stimuli. A native speaker of English was recorded saying 160 CVCVC words. Each word began with one of 80 CVC stems, twenty each with the vowels /i/, /e/, /u/ and /o/ and random consonants. Each stem was recorded with both variants, or allomorphs, of a suffix, [it] and [ut]. Thus, half the words were front-harmonic (e.g., pel-it, bis-it) and half were front-disharmonic (e.g., pel-ut, bis-ut).

Procedure. The procedure followed a modified artificial grammar paradigm. Participants were assigned to one of two conditions: the harmonic condition or the height-front dependency condition, which is unattested. In both conditions, participants were exposed in training to 40 words from the language they were learning. In the harmonic condition, 40 harmonic words were selected. In the height-front dependency condition, words were selected such that mid-vowel stems received the front vowel suffix (e.g., pel-it, bod-it) and high-vowel stems received the back-vowel suffix (e.g., bis-ut, tug-ut). This rule was chosen arbitrarily from the space of possible languages to test the hypothesis that vowel harmony would have a learnability bias over other patterns.

Participants were familiarized with the words in the same way regardless of condition. They were given alternating blocks of passive listening and blocks in which for each trial, two words were played and they were required to choose which word they had previously heard. In the forced choice trials, the choice was between a word that had been played in the passive listening section and a word with the same prefix and the alternate allomorph. A total of five blocks of 40 trials each were included in training: three passive listening blocks with a forced choice block in between each.

Following the training trials, participants completed one block of 80 test trials. On each test trial, participants were asked to choose which of two words they thought was from the language they had learned in the training trials. In each trial, the two words both had the same stem and differed in the suffix. 40 of the test trials included words from training, and 40 were generalization trials involving novel words.

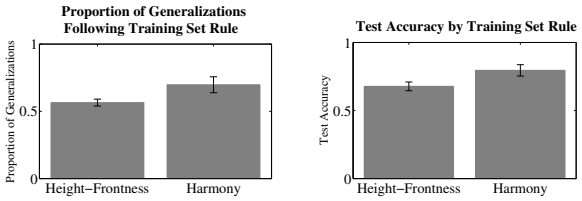


Figure 3: Results for harmonic versus height-frontness rule conditions. By condition, there are significant differences in the proportion of generalizations following the rule (0.70 for harmony rule versus 0.57 for height-frontness rule, $t(38) = 2.05, p < 0.05$; left) and in test accuracy (0.80 for harmony rule versus 0.68 for height-frontness rule, $t(38) = 2.23, p < 0.05$; right).

5.2 Results

As shown in Figure 3, we found a learnability bias for the harmonic language. Learners had significantly greater accuracy in test when they learned the vowel harmonic language than when they learned the height-front dependency language (80% correct for learners of the harmony rule versus 68% correct for the height-frontness rule, $t(38) = 2.23, p < 0.05$). Additionally, 70% of generalizations made by learners in the harmony rule condition followed the harmonic rule while only 57% of generalizations made by learners in the height-front dependency condition followed the height-frontness rule ($t(38) = 2.05, p < 0.05$).[2] The result of these two phenomena was that the final languages produced by the learners in the harmony condition had a greater prevalence of harmonic words than the final languages of learners in the height-frontness dependency had of adhering words.

These results establish that the probability of transitioning from a harmonic language to another language with a high proportion of harmonic words is higher than the probability of transitioning from a height-front dependency language to another language with a high proportion of adhering words. In

[2]For the second experiment, participants who had low accuracy ($< 62.5\%$ of previously heard words chosen in test as "from the language") were excluded. Performing this exclusion in this experiment preserves the same results: Mean accuracy of 87% for the harmonic condition versus 73% for the height-front dependency condition ($t(28) = 2.74, p < 0.025$), and 77% mean proportion of generalizations following the rule for the harmonic condition versus 58% for the height-front dependency condition ($t(28) = 2.43, p < 0.025$). This exclusion criterion resulted in removing five participants from each condition.

terms of the transition matrix, this corresponds to $q_{\ell_{\text{harm}},\ell_{\text{harm}}} > q_{\ell_{\text{h-f}},\ell_{\text{h-f}}}$, where ℓ_{harm} is the set of languages with a high proportion of harmonic words and $\ell_{\text{h-f}}$ is the set of languages with a high proportion of words that follow the height-frontness rule. In other words, the harmonic language is easier to learn than the height-front dependency language.

6 Experiment 2: Language Transmission

6.1 Methods

Participants. There were a total of 104 participants who received either monetary compensation or course credit for their participation. All were native speakers of English.

Stimuli. The same stimuli were used as in Experiment 1.

Procedure. The procedure for this experiment was similar to Experiment 1, but the way that words were chosen for training differed. For the first subject in each chain, a total of 40 prefixes were selected at random, and based on the starting condition of the chain, the allophone for each prefix was selected. For example, for the 50% harmonic starting condition, 40 prefixes were chosen and of those prefixes, half were chosen to have the appropriate allophone to make the word harmonic and half were chosen to have the allophone to make the word nonharmonic. For subsequent subjects in each chain, 40 words were chosen at random from those words which the previous subject had said was in the language. In order to exclude subjects who had not actually learned the language in training, subjects were not included in the chain if their accuracy in test on previously seen words was below 62.5%; this is the lowest level of accuracy that is significantly different (binomial test, $p < 0.05$) from chance guessing. Chains were started at 100%, 75%, 50%, 25%, and 0% harmonic. One chain with 10 subjects was run for each starting point except for 100%. Four chains of 10 subjects each were run at this starting point as this is the point of most interest: given a learnability bias, does the percentage of harmonic words in a language remain consistently large?

6.2 Results

While Experiment 1 showed a learnability bias for the harmonic language over an arbitrarily chosen language, the iterated learning chains in Experiment 2 did not favor the harmonic language. As shown in Figure 4, all chains tended toward languages with approximately 50% harmonic words, and after several generations, the chains that began with 100% harmonic words did not differ significantly from the other chains. There is also no difference in accuracy on the harmonic items over time, as shown in Figure 5. This is empirical evidence that the pattern shown in simulation can also occur with human learners: One language is more accurately transmitted than others, but due to the large number of other possible languages, this language does not predominate after many transmissions.

7 General Discussion

In this paper, we formalized language transmission using a linear model in order to examine whether a learnability bias for some property of language is sufficient for that property to become prevalent in human languages. We showed two ways in which a learnability bias for a property can exist but not cause that property to become prevalent. First, using a mathematical analysis, we showed that this can occur when transmission failures favor languages other than those that have greater learnability. This illustrates the importance of considering the entire transmission matrix, not just the probabilities of accurate transmissions that are considered when establishing a learnability bias.

Second, we showed that it is possible for the sheer number of other possible languages to overwhelm greater learnability for a particular language. We then illustrated that this second scenario might lead to incorrect predictions in an experimental context. In artificial language experiments, greater learnability is often established by comparing the accuracy of transmission for a language with the property of interest to an arbitrary language. However, in our experiment, we established such a learnability bias for vowel harmony, but this did not result in vowel harmony being maintained after many instances of transmission. This result seems to be due to the fact that numerous languages other than harmonic languages were possible, so learners tended to learn one of these many other languages.

One limitation of our analysis is the use of the

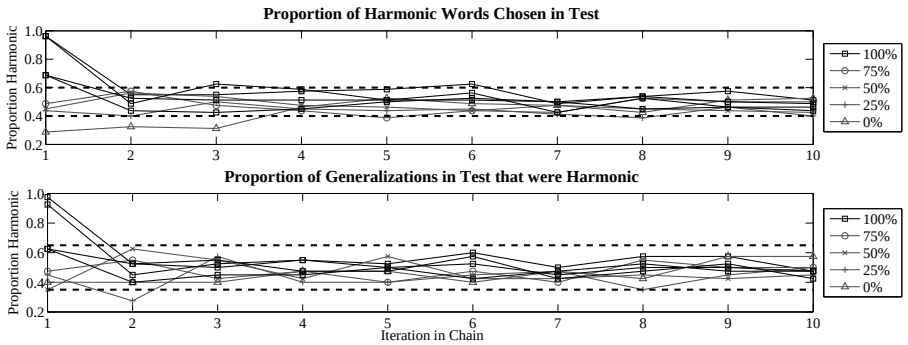


Figure 4: Iterated learning chain results. Dotted lines show the two-tailed 95% confidence interval for chance responding; confidence intervals differ between the two graphs because there are 40 opportunities to generalize versus 80 opportunities to choose harmonic words.

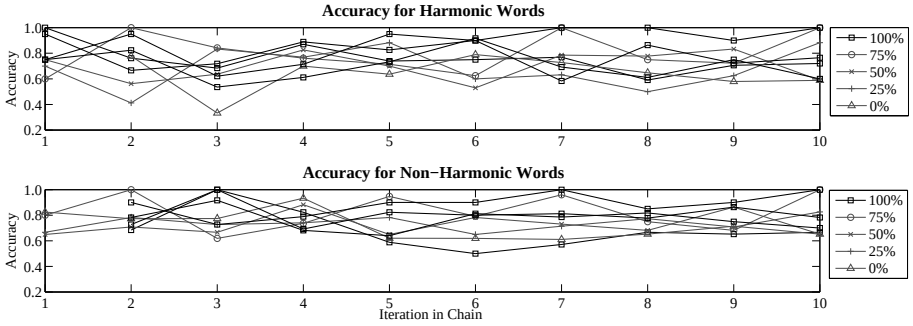


Figure 5: Accuracy on harmonic versus non-harmonic words by iteration. Overall, there is no difference in accuracy.

simple linear transmission model, in which each learner learns from one member of the previous generation. It is easy to imagine variants on this model that make more realistic assumptions about cultural transmission of languages. However, we suspect that these more complex models would not alter the conclusions that we have drawn here. For example, learning from multiple members of the previous generation tends to dilute the effects of learnability on the languages produced by a population (Smith, 2009; Burkett & Griffiths, 2010).

Overall, the result of a more complicated relationship between learnability biases and linguistic universals is congruent with the evidence that all languages do not exhibit all properties for which learnability biases have been found. Indeed, in historical linguistics, the general principle is one of language divergence, rather than convergence on some universal language (e.g., Greenberg, 1971). Given this relationship, one must rethink using experimental evidence for particular learnability biases to explain linguistic tendencies. Instead, one must either estimate all of the values in the transmission matrix, or actually simulate the process of multiple transmissions in the lab to establish whether a particular property with a learnability bias is actually maintained over many generations. While this process is dependent on assuming a particular model of how transmission occurs in populations, such as the linear iterated learning paradigm we used in our experiments, it provides a way of understanding what mutations are likely to occur and of exploring the long term trends that result from particular learnability biases. As we showed for vowel harmony, long term trends may not match what one predicted based on a learnability bias. Given such a result, one must look to factors other than the learnability bias to explain why a property is common across languages.

Acknowledgements. This work was supported by an NSF Graduate Research Fellowship to ANR, grant number BCS-0704034 from the NSF to TLG, and grant number T32 NS047987 from the NIH to ME.

References

Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in turkish laryngeal alternations. *Language*, *87*(1), 84-125.

Burke, C. J., & Rosenblatt, M. (1958). A Markovian function of a Markov chain. *The Annals of Mathematical Statistics*, *29*(4), 1112–1122.

Burkett, D., & Griffiths, T. (2010). Iterated learning of multiple languages from multiple teachers. In *The Evolution of Language: Proceedings of the 8th International Conference (EVOLANG8).*

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vannemann (Eds.), *Syntax: An international handbook of contemporary research* (pp. 506–569). Berlin: Walter de Gruyter.

Comrie, B. (1981). *Language universals and linguistic typology*. Chicago: University of Chicago Press.

Croft, W. (2002). *Typology and universals*. Cambridge University Press.

Finley, S., & Badecker, W. (2007). Towards a substantively biased theory of learning. *Berkeley Linguistics Society*, *33*.

Finley, S., & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, *61*, 423–437.

Greenberg, J. (Ed.). (1963). *Universals of language*. Cambridge, MA: MIT Press.

Greenberg, J. (1971). *Language, culture, and communication*. Stanford: Stanford University Press.

Griffiths, T. L., & Kalish, M. L. (2007). A Bayesian view of language evolution by iterated learning. *Cognitive Science*, *31*, 441-480.

Kemeny, J., & Snell, J. (1960). *Finite markov chains*. Princeton, NJ: van Nostrand.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, *5*, 102-110.

Komarova, N. L., & Nowak, M. A. (2003). Language dynamics in finite populations. *Journal of Theoretical Biology*, *221*, 445-457.

Labov, W. (2001). *Principles of linguistic change. Volume II: Social Factors*. Blackwell.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*(1), 83–127.

Rafferty, A. N., Griffiths, T. L., & Klein, D. (2009). Convergence bounds for language evolution by iterated learning. *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society*.

Schuster, P., & Sigmund, K. (1983). Replicator dynamics. *Journal of Theoretical Biology*, *100*(3), 533 - 538.

Slobin, D. (1973). Cognitive prerequisites for the acquisition of grammar. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (pp. 173–208).

Smith, K. (2009). Iterated learning in populations of Bayesian agents. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

van der Hulst, H., & van de Weijer, J. (1995). Vowel harmony. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 495–534). Blackwell.

Wilson, C. (2003). Experimental investigation of phonological naturalness. *Proceedings of the 22nd West Coast Conference on Formal Linguistics*.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945–982.