NORTHWESTERN UNIVERSITY

Role of Prior Knowledge in Adult Learning of Second-Order Phonotactic Generalizations for Speech Production

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Linguistics

By

Nicole Irene Mirea

EVANSTON, ILLINOIS

December 2022

Abstract

Phonotactic patterns are generalizations that govern the order of consonants and vowels, within words and syllables. Certain second-order phonotactic patterns—those that relate multiple sounds within a syllable, such as "if the vowel is [ɪ], then [s] can only appear at the end of the syllable"—require a period of sleep-based consolidation before adults can internalize them such that their speech errors conform to these generalizations in a tongue-twister task (Gaskell et al., 2014). The present work examines the role of prior knowledge in consolidating these second-order phonotactics for speech production, testing the hypothesis that new phonotactic knowledge which is similar to previously-acquired knowledge is easier to consolidate during sleep. We do this by analyzing errors made in a tongue-twister task (Chapter 2). We operationalize "similar prior knowledge" in terms of the syllable positions to which the consonants are constrained, and consider two sources of prior knowledge in these speech experiments: participants' native language (English; Experiment 1) and pre-trained restrictions learned through tongue-twisters produced in the first experimental phase (Experiment 2). In a follow-up experiment in a non-linguistic sequential action domain (keypress sequence production; Chapter 3), we test whether the results we found in Experiment 1 were specific to speech production, or due to more general properties of sequential action. Our results indicate that global (first-order) and local (second-order) sequential patterns are simultaneously represented in production, and that this may be a necessary property of the schema for structured sequential action, in both speech and keypress sequence production. We found evidence of transfer from phonotactic restrictions that had been previously learned in a similar environmental context, supporting an incremental account of phonotactic learning (Anderson et al., 2019) as well as a view of phonotactic learning in which restrictions are stored together with their

contexts of acquisition (Dell et al., 2019). Our results also suggest that consolidation's main function in phonotactic learning for speech production is to resolve interference among multiple conflicting mappings between consonants and their syllable positions. Moreover, we found evidence that participants' native language (English) may have affected their speech errors. Overall, this research demonstrates that prior knowledge's complex role in sequential pattern learning, which depends not only on similarity to new information and opportunity for consolidation, but also on the environment in which that prior knowledge was acquired.

Acknowledgments

Like many PhD students, I have been keeping a document called "Acknowledgments" handy for occasions when my enthusiasm for my dissertation project has been especially low. Through the setbacks and dead ends that characterize an endeavor like this, I have taken comfort in having the assistance and encouragement of so many wonderful people in my life. I am extraordinarily lucky to have benefitted from their support, and I am happy that I get to memorialize it here.

First and foremost, I thank my husband Dan Turner, on whose love, understanding, and patience I rely daily: for all the times you walked Frieda without me, for all the snacks you brought me while I was "just finishing up," and for all the pep talks and complaint sessions. I thank my other friends for keeping me grounded, hopeful, and somewhat sane: Maddie Forrest, Jean Delkhaste, Arielle Elliott, Rox Sayde, Tommy Denby, Lisa Cox, Haley Farkas, Shawn Foster, Adam Goodkind, the entire NLP SPC (current and former), and many others.

Throughout my graduate career, I've benefited from the solid foundation in linguistics and cognitive science that my mentors have given me. At UCLA, Robert Daland introduced me to the puzzle of phonology, and was the first person to ask me if I'd ever thought about going to grad school. Doe Buchli introduced me to the idea that learning could involve offline processes. Everett Mettler introduced me to the highs and lows of experimental research on learning and forgetting. In my first few years at Northwestern, Klinton Bicknell introduced me to computational modeling of behavioral data. And Matt Goldrick, of course, introduced me to the usefulness of errors (including my own), as well as the value of leadership through service.

Beyond my thesis committee, there are several other people who have had a more direct contribution to this project. My peers in Soundlab (Oriana Kilbourn-Ceron, Kasia Hitzcenko, Maria Gaviño, Lisa Cox, and Qile Chen) provided early feedback that improved my ideas significantly. Kasia Hitzcenko deserves special mention for being my second transcriber for the remote speech recording pilot in Spring 2020; without that pilot data, all this would have remained theoretical for at least another two years. Our department's software developer, Chun-Liang Chan, wrote the module that allowed us to record sound remotely. The contribution of undergraduate research assistants Victoria Ozymko and Camila Hirani truly cannot be understated: Victoria spent at least 150 hours tirelessly transcribing tongue-twisters as Transcriber #2—an *enormous* amount of work that I could barely keep up with as Transcriber #1—and Camila kickstarted data collection during a global pandemic, no less!

Much of this research can be classified as "Your tax dollars at work." It was chiefly funded by the National Science Foundation (NSF)'s Doctoral Dissertation Improvement Grant (BCS 2116802), and I was an NSF Graduate Research Fellow while working on it. Participant recruitment was facilitated by Northwestern's Communication Research Registry (supported by NIH grants UL1TR001422 and P30DC012035). This work was also funded by a Graduate Research Grant from the Graduate School of Northwestern University. I am fortunate to live in a time and a place where such things are possible.

And I wouldn't—if not for the concerted, sustained, and immense efforts of my parents, Marina and Adrian Mirea. As I approach the age at which my mother left her country for this strange land, I am increasingly aware of what a struggle that was; thanks to my parents' sacrifices, I hope I never need to make that choice for myself. I am especially grateful that I was always encouraged to ask questions, even when the answers were not at hand. My grandmother, Valeria Danciu, taught me my mother tongue, my times tables, and so much more; this work is dedicated to her.

Bunicii mele, care m-a învățat limba maternă și tabla înmulțirii, între multe altele

Table of Contents

List of Tables

List of Figures

Introduction

Language learning is a dynamic process: the current state of the linguistic system affects its future state. An adult native Polish speaker and an adult native Spanish speaker can take the same English course, but it is likely that the native Spanish speaker will have more difficulty pronouncing word-initial /s/-consonant clusters (as in the word <stop>) after an equivalent amount of practice, since these clusters are prohibited in their native language (Harris, 1983, p. 14)—and indeed, several studies have documented L2 Spanish-English learners' tendency to modify these clusters with an prothetic vowel (e.g., stop → "estop"; see Carlisle, 1988 for a review and additional supporting data), even after at least one year of English experience (Carlisle & Cutillas Espinoza, 2014). By contrast, the native Polish speaker has more experience with these clusters (found in Polish words like <styl> /stɨl/ "style," <stopka> /ˈstɔp.ka/ "footnote," and others) and therefore does not need to modify them when producing English words (Hodne, 1985).

One mechanism through which prior knowledge may influence learning is consolidation—an improvement in what learners can recall across a period of sleep or resting while awake. According to McClelland et al. (1995), consolidation's purpose is to integrate new knowledge with old, such that old knowledge is not overwritten by new. Lewis and Durrant (2011) additionally propose that this integration helps learners pick up on patterns by replaying both new and old memories at the same time. Based on these theories, we propose that certain prior knowledge may have differential effects on participants' knowledge state, and that these differential effects may only emerge after a consolidation period. We hypothesize that new knowledge which is similar to previously acquired knowledge should benefit more from consolidation than dissimilar new knowledge, since dissimilar new knowledge may require additional time or resources before it can be integrated with prior knowledge during consolidation.

Here, we test this hypothesis in the realm of phonotactic pattern learning for speech production. Phonotactic patterns are generalizations that govern the order of consonants and vowels, within words and syllables. These generalizations vary from language to language, as in the Polish vs. Spanish example above. Several previous studies have found consolidation effects in learning novel phonotactic patterns, as measured through speech errors' tendency to follow phonotactic restrictions governing stimuli of a tongue-twister task (Warker & Dell, 2006; see Anderson & Dell, 2018 for meta-analysis). A consolidation period appears to be required before adults' speech errors in this task can reflect novel second-order phonotactic patterns (Gaskell et al., 2014)—i.e., those that relate multiple sounds within a syllable, such as "if the vowel is [ɪ], then [s] can only appear at the end of the syllable."[1] Such patterns are common across the world's languages. The present work clarifies the role of prior knowledge in consolidating these second-order phonotactics.

1.1 Approach

To test our hypothesis, we collected data from three experiments manipulating the relationship between prior knowledge and to-be-learned sequential pattern generalizations. The first two of these (described in Chapter 2) were speech production experiments, in which participants recited tongue-twisters conforming to a set of phonotactic restrictions. In the third experiment (described in Chapter 3), which was a follow-up in a non-linguistic modality, participants keyed in sequences of keypresses that were constructed according to a set of rules akin to the phonotactic restrictions from the speech experiments.

[1] This is in contrast to first-order phonotactic patterns, which can affect speech errors before a consolidation period (Warker & Dell, 2006).

1.1.1 Experiments and Predictions

In Experiment 1, participants learned phonotactic restrictions that either resembled or differed from structural associations found in our learners' native language (English). Per our hypothesis, we expected structurally similar phonotactics to benefit more from consolidation, since prior knowledge could be leveraged more readily. We operationalized this "structural similarity" in terms of syllable positions that were constrained by the vowel in each syllable of the tongue-twisters. In English, syllable codas are more predictable from vowel identity than syllable onsets (Kessler & Treiman, 1997), and so stimuli seen by participants in the *English-similar* condition conformed to phonotactic restrictions that constrained consonants to coda, based on the identity of the vowel (e.g., "if the vowel is [ɪ], then [s] can only appear in coda"). By contrast, in the *English-dissimilar* condition, participants saw stimuli conforming to phonotactic restrictions that constrained consonants to onset (e.g., "if the vowel is [ɪ], then [s] can only appear in onset").

In Experiment 2, the source of prior knowledge was a pre-training phase. This design gave us more control over the specific prior knowledge our participants had coming into the task, as well as the context in which participants acquired that prior knowledge. Phonotactic restrictions learned in the main experimental phase were either similar or different from the pre-training phase, in terms of the syllable positions to which consonants were restricted. As in Experiment 1, we predicted that if consolidation recruits prior knowledge for phonotactic learning, participants in the *training-similar* condition would be more efficient at consolidating the new phonotactic generalizations, with more improvement after a consolidation opportunity than their *training-different* counterparts. This experiment additionally allowed us to explore the degree of transfer between already-consolidated and novel second-order phonotactic patterns.

To test whether the results of Experiment 1 were due to linguistic experience and not differences in planning processes between conditions, we replicated Experiment 1 in a sequential action domain where participants do not have structured prior experience: button-pressing (Chapter 3). We expected less of a difference between conditions in this experiment, compared to in Experiment 1. The critical manipulation and dependent measure were identical to those in the speech experiments, but participants keyed in sequences of button-presses on specially modified keyboards instead of producing syllables.

1.1.2 Measuring Knowledge through Errors in Production Tasks

In all experiments, participants were asked to produce sequences to the beat of a metronome. Each trial began with one repetition of the stimulus sequence at a slow pace (1 beat per second), followed by three repetitions at a fast pace (2.53 beats per second); only errors in the fast repetitions were included in our analysis. Each stimulus sequence comprised multiple smaller subsequence units (four syllables in the speech experiments, three panes in the button-pressing experiment). In all experiments, these subsequences contained exactly three "atomic" response units: consonants and vowels in the speech experiments and keypress targets in the button-pressing experiment. Syllables in the speech experiments always had the form consonant-vowel-consonant. Similarly, each pane in the button-pressing experiment displayed a sequence of keypress targets in which the first and third targets were always finger keys (index, middle, or ring on either hand) and the second target was always a thumb key. In the speech experiments, the vowel was always the same (either /æ/ or /ɪ/) in all four syllables of each stimulus, and no consonants were repeated within the stimulus. Similarly, in the button-pressing experiment, the thumb key was always the same (either left or right) in all three panes of each stimulus, and no finger key targets were repeated within the stimulus.

All stimuli contained three types[2] of targets (consonants or finger keypresses), categorized according to the positions in which they could occur within subsequences (onset or coda in speech experiments, first or third in button-pressing experiment):

1. *Unrestricted* targets could appear in any position, in any stimulus.
2. *Restricted-constrained* targets were restricted to a particular position in the context of the vowel or thumb in the current stimulus.
3. *Restricted-free* targets were restricted to a particular position elsewhere in the experiment, but not in the context of the vowel or thumb in the current stimulus.

We inferred the state of participants' knowledge on each day by analyzing their errors. More specifically, we measured the extent to which participants' errors conformed to the generalizations in question by comparing the rates at which the aforementioned three target types preserved their syllable position in participants' errors. Knowledge of the phonotactic restrictions that generated the stimuli was indexed by how much more often errors on *restricted-constrained* targets preserved position, compared to errors on both *restricted-free* and *unrestricted* targets.

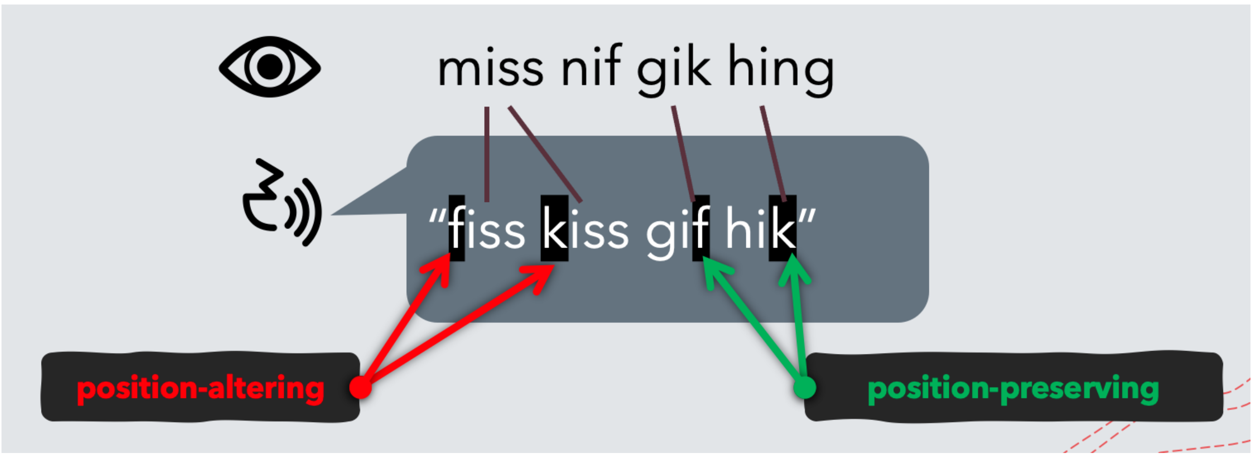
To assess this, we first coded each error as position-preserving if the produced consonant or keypress appeared in the same position as it did in the stimulus, and position-altering otherwise. An example of this coding is depicted in Figure 1. To test whether our predictions were supported by the data, we then used each error as an observation in a logistic mixed-effects model that included target type (as well as other relevant factors) as predictors.

---

[2] Speech experiment stimuli additionally included two language-restricted consonants, /h/ and /ŋ/, which were not included in our main analyses..

Figure 1

Coding Errors for Position-Preservation



Note. Gray box depicts sample production by participant, in response to the tongue-twister stimulus above it. Purple lines indicate alignment to target syllables, calculated according to the algorithm described in Section 2.3.2.2.1. Both /f/ and /k/ appear in coda in the stimulus, so the errors in which these consonants are produced in onset position are classified as position-altering. Section 2.3.2.2.1 describes how this classification was used in our statistical analysis.

1.1.3 Testing for Effects of Slow-Wave Sleep

To assess whether prior knowledge's influence on learning took place during sleep-based consolidation (as opposed to some other stage of acquisition), participants took part in two sessions per experiment, with an opportunity for consolidation between each session. Slow-wave sleep, a stage of deep sleep characterized by long slow delta waves and slow or absent eye movement, has been implicated in second-order phonotactic learning (Gaskell et al., 2014). Given this, we calculated the proportion of slow-wave sleep between nights of the study for each participant, using data collected from wrist-worn activity monitors. Based on Gaskell et al. (2014)'s findings, we expected individual differences in slow-wave sleep proportion to correlate with the magnitude of improvement in phonotactic learning across days. We additionally reasoned that, if consolidation was implicated in prior knowledge's effect on learning, this correlation should be stronger for participants in the *similar* condition of each experiment.

1.2 Findings and Significance

Overall, we found that the function of consolidation in second-order phonotactic learning depended on the nature of prior knowledge and how it was acquired. In the speech experiments, consolidation enabled transfer from prior phonotactic knowledge, but only when that knowledge was acquired in a similar environment as the task itself (i.e., pre-training). In contrast to previous second-order phonotactic learning studies where mappings between restricted consonants and the positions they appeared in interfered with each other, we found evidence of phonotactic learning on Day 1 of each experiment, suggesting that consolidation's role in previous studies had been to resolve interference between conflicting mappings. Participants tended to keep syllable rimes intact in speech errors; we failed to replicate this in the button-pressing task, suggesting that this tendency arose from distributional characteristics of participants' native language (English). We additionally found evidence that participants generalized second-order phonotactic restrictions to context-independent, first-order constraints that bound restricted consonants to a particular syllable position more often than chance, even when not constrained by immediate vowel context.

Our results support a contextually compartmentalized view of phonotactics for speech production, in which representations of phonotactic generalizations are tied to the environment they are learned in, as in the model proposed by Dell et al. (2019). Learners may represent second-order phonotactic generalizations as confluences of context-dependent and context-independent positional restrictions. This research additionally indicates that consolidation's main function in phonotactic learning for speech production is to resolve interference among multiple conflicting mappings between consonants and their syllable positions. This work further clarifies how the neural mechanisms underlying speaking adapt to new input over time, integrating the study of linguistic knowledge with work in the cognitive neurosciences.

Learning Second-Order Phonotactics in Speech Production

2.1 Introduction

Although learning a new language as an adult is difficult (Newport, 1990), adults may be able to draw upon their prior knowledge to help them fluently pronounce words and sentences in a new language. The mature speech production system can adapt such that speech conforms to novel *phonotactic restrictions*—i.e., restrictions governing the ordering of sounds within words and syllables—that are not present in the learner's native language, particularly if the novel phonotactic restrictions generate words and phrases that are legal in the learner's native language (Dell et al., 2000). Language learning is a dynamic process, so the current state of the linguistic system affects the future state of the system; however, the extent and nature of prior knowledge's influence on phonotactic learning have not yet been ascertained. In this multiple-day training study, we examine effects of prior phonotactic knowledge from two sources: participants' native language (English) and pre-training in the experimental context. We additionally investigate sleep-based consolidation's role in moderating the relationship between old and new phonotactic knowledge.

2.2 Learning Phonotactic Constraints for Speech Production

2.2.1 *Speech Errors as a Window into Phonotactic Knowledge*

Wells (1951/1973) first noted that "a slip of the tongue is practically always a phonetically possible noise," remarking that this "phonetic possibility" includes the dispreference of "English sounds… in un-English combinations" (p. 86). For example, native English speakers are more likely to mispronounce the target phrase "sleepy vole" as "<u>fl</u>eepy sole" than "<u>vl</u>eepy sole," even though neither "fleepy" nor "vleepy" are words of English—simply because the onset "fl" is more phonotactically probable in English than "vl."

This observation, along with other converging evidence from naturally-produced speech error data (Boomer & Laver, 1968; Fromkin, 1971; Motley, 1973), led Motley and Baars (1975) to verify this tendency experimentally by using a tongue-twister task to induce speech errors in the laboratory.

Since then, several other researchers have leveraged this property (dubbed the *phonotactic regularity effect* by Dell et al., 1993) to infer the current state of speakers' phonotactic knowledge available for speech production by working backwards from their speech errors. The logic of this is as follows: if knowledge of a phonotactic restriction is used to guide speech production, then participants' errors should follow a phonotactic restriction more often than expected by chance. Therefore, by calculating the proportion of speech errors that follow a phonotactic restriction and controlling for additional factors that influence speech production, a researcher can obtain a metric that corresponds to the strength of knowledge regarding a phonotactic restriction that is available for speech production.

Speakers can make use of knowledge from multiple sources. Notably, Dell et al. (2000) found that speakers' errors are not only sensitive to statistical regularities found in their native language, but also reflect statistical regularities from recently-presented stimuli. In this study, native English speakers recited tongue-twisters consisting of consonant-vowel-consonant "words" in which certain consonants were restricted to particular syllable positions by novel phonotactic restrictions. When aggregated across all four days of the study, all groups of participants' speech errors reflected the statistical regularities in the stimuli, as well as phonotactic restrictions found in English. The production system, therefore, adapts to reflect recent experience, and this adaptation can be quantified and detected by analyzing speech errors.

2.2.2 *Consolidation Effects in Phonotactic Learning*

Although all groups of participants in Dell et al. (2000) did eventually learn the phonotactic restrictions used to create the stimuli, regardless of the phonotactic restriction in question, not all phonotactic restrictions are learned equally quickly. Several follow-up studies using the same tongue-twister paradigm have failed to detect effects of novel phonotactic restrictions on Day 1 in the speech errors of participants exposed to phonotactic regularities in which the position of consonants depends on the identity of the local vowel (i.e., *second-order phonotactic restrictions*), despite finding Day 1 adaptation for speakers exposed to global, vowel-independent constraints (Warker & Dell, 2006; see Anderson & Dell, 2018 for meta-analysis). Results from other recent studies (Smalle et al., 2017, 2021; Smalle & Szmalec, 2021) which find Day 1 adaptation for second-order phonotactic constraints in populations with reduced attentional control (children, older adults, and cognitively-depleted adults) suggest that consolidation may serve to overcome difficulties with pattern generalization that are caused by the failure of participants with heightened attentional control to consider the context necessary to learn second-order constraints.

Gaskell et al. (2014)'s nap study established that it was sleep, and not the mere passage of time, that enabled participants' speech errors to reflect their recent experience. In this study, each participant performed the tongue-twister task for two sessions, spaced 110 minutes apart, with stimuli in both sessions following the same second-order phonotactic restriction. In the interval between sessions, the researchers randomly assigned each participant to either nap for 90 minutes or watch a video with minimal dialogue. The rate at which speech errors followed the phonotactic restriction increased significantly more for participants who slept between sessions, compared to participants who remained awake. Gaskell et al. (2014) additionally recorded electroencephalography (EEG) data during participants' naps and found positive correlations between second-order phonotactic learning in speech production and characteristics related to the depth of slow-wave sleep: namely, duration of this sleep stage and amplitude of the low-frequency neural oscillations (delta waves) that characterize it. Other research has also implicated slow-wave sleep in sleep-based consolidation—particularly in the development of abstraction over hidden patterns by transferring information between the neocortex and hippocampus (see Diekelmann & Born, 2010, for a review). This accumulation of evidence suggests a critical role for sleep-based consolidation in the learning of second-order phonotactic restrictions.

What is the causal mechanism linking sleep-based consolidation to adaptation in speech production? According to Stickgold & Walker (2013)'s theory of consolidation, sleep-based consolidation serves to increase the usefulness of knowledge for future actions. In this case, the "future action" is better performance on the tongue-twister task, with the speaker's memories of the experimental task integrating with existing phonotactic knowledge. In a similar vein, Dell et al. (2021) proposed that learners accumulate knowledge of second-order phonotactic constraints during first exposure, and consolidation converts it into a form that can subsequently influence speakers' productions—perhaps by foregrounding information previously ignored due to attentional control, as Smalle et al. (2021) have suggested. In Warker & Dell (2006)'s connectionist model of this process, this is achieved through the activation of "hidden units" that code for context dependencies within the rule—*e.g.* [s] in the context of [ɪ]—during the consolidation period. According to Dell et al. (2021), "[c]onsolidation promotes the separation [between the experimental context and English], making it so that the new rule, rather than contrary English rules, is in force in the experimental context during the second session" (p. 20).

This “separation-promotion” is one example of how consolidation mediates the relationship between existing and newly-learned phonotactic knowledge.

2.2.3 *Current Study: A Role for Prior Knowledge?*

In this study, we test whether consolidation may have another, more selective function in phonotactic learning. According to Lewis and Durrant (2011)’s theory of schema consolidation and generalization, newly-formed memories and old related ones are reactivated during slow-wave sleep through neuronal replay—an offline process in which neurons activate in patterns that are similar or even identical to their activations during online practice (Genzel & Robertson, 2015). This replay selectively strengthens the overlapping parts of memories and promotes abstraction over their commonalities. We hypothesize that, in the context of phonotactic learning, congruence of old and new phonotactic constraints can create this overlap, leading to more successful integration of novel phonotactic constraints that are similar to previously-learned ones. Additionally, we propose that phonotactic dependencies can be gradiently accessible, such that some second-order phonotactic dependencies may benefit more from consolidation than others, depending on whether the hidden units governing these dependencies were already active in the learner’s productive repertoire.

If our hypothesis is correct, we expect phonotactic patterns that are more structurally similar to previously-learned phonotactic constraints to benefit more from consolidation. We consider two sources of prior knowledge in this study: participants’ native language (Experiment 1) and pre-training in the context of the experiment (Experiment 2). In designing the experiments in this study, we operationalize this “structural similarity” with respect to Warker & Dell (2006)’s model of speech production, assuming two constraints are similar if they share the same relationship between syllable positions—*i.e.*, if they activate the same hidden unit. In Experiment

1, we leverage the fact that in English, vowel identity predicts syllable codas more reliably than it does syllable onsets (Kessler & Treiman, 1997). Therefore, participants in our *language-similar* condition recited words of an artificial language containing dependencies between vowels and coda consonants—*e.g.* "If /ɪ/ is the vowel, then /s/ can only be a coda." In contrast, we will expose participants in the *language-dissimilar* condition to syllables with dependencies between vowel and syllable onsets—*e.g.* "If /ɪ/ is the vowel, then /s/ can only be an onset." Participants from Experiment 1's *language-dissimilar* condition also took part in Experiment 2, where their experience from Experiment 1 was used as pre-training. In Experiment 2, participants in the *training-similar* condition recited tongue-twisters with a similar vowel-onset dependency, and those in the *training-dissimilar* condition recited tongue-twisters containing a vowel-coda dependency. In both experiments, our hypothesis predicted that the increase in likelihood of errors of participants in the *similar* condition to follow the learned phonotactic restrictions would be greater than the increase in this likelihood for participants in the *dissimilar* condition.

2.3 Experiment 1: Prior Knowledge from English

2.3.1 *Method*

2.3.1.1 Participants. We recruited 115 adult native English speakers with no self-reported history of speech, language, hearing, or sleep disorder through flyers posted on the Northwestern University campus, as well as via Northwestern University's Communications Research Registry. Each participant completed a screening questionnaire ahead of participation to establish that they met the aforementioned demographic criteria, as well as to ensure that they had access to the equipment required to complete the study: a computer, headphones, and a non-Bluetooth microphone. This screening survey also contained questions about availability, and participants were only eligible if they could be available for 30 minutes at the same time of day for both two weekdays in a row and for four weekdays in a row. The purpose of these questions was to minimize confounding differences between participants in the language-similar condition (who only took part in the study for two days) and the language-dissimilar condition (who went on to take part in Experiment 2 after Experiment 1, for a total of four days of participation).

All participants provided informed consent before taking part in the study. Participants were compensated for completing sessions on two schedules, depending on condition. Participants in the language-similar (2-day) condition were compensated a total of $20 for their participation ($7 for Day 1 + $13 for Day 2). In order to incentivize return for the two subsequent Experiment 2 sessions, participants in the language-dissimilar condition were paid a total of $16 for completing Experiment 1 ($7 for Day 1 + $9 for Day 2), and an additional $24 for completing Experiment 2 ($11 for Day 1 + $13 for Day 2).

A total of 47 participants were excluded due to experiment software malfunction (N = 13), poor audio quality (N = 11), attrition after Day 1 (N = 6), not making at least one error on the relevant syllable positions for analysis (see "Dependent Measure," Section 2.3.2.2.1) in both sessions (N = 6), failing to complete at least one session (N = 4), experimenter error in verifying eligibility requirements (N = 3), failing to sleep for at least 3 hours between sessions (N = 3), taking more than one hour to complete either session (N = 2), and/or audible distractions (N = 2), for a final sample size of 68 participants.

The target number of participants recruited was determined by a power analysis. To determine the power of this experiment to detect a significant effect, we ran a Monte Carlo simulation to generate the data we expected to collect, given a range of sample sizes. Using effect size estimates from previous studies using this paradigm (Gaskell et al., 2014; Goldrick & Larson, 2008), we constructed each dataset such that simulated participants in the ‘similar’ condition were, on average, 1.49 times more likely to preserve syllable position of targeted consonants in errors on Day 2 compared to simulated participants in the ‘dissimilar’ condition. We generated 1000 datasets for each sample size tested, and ran the planned mixed-effects logistic regression from our pre-registration (Mirea & Goldrick, 2021) on each dataset. We began our testing at a sample size of 24 participants in each experimental condition, because we wanted to be able to recover the second-order phonotactic learning effect in both experimental conditions, and the largest study to do so to date (Warker & Dell, 2015) used 24 participants. At the minimum tested sample size of 24 participants per condition, $1 - \beta = 0.76$.[3]

2.3.1.2 Materials. All stimuli were tongue-twisters consisting of four syllables, each with a consonant-vowel-consonant structure[4]. Tongue-twisters were presented with a space between each syllable. Within each tongue-twister, the same vowel was used in all four syllables to encourage errors (Wilshire, 1999), but, as in Dell et al. (2000) and other experiments using the tongue-twister paradigm, each consonant appeared exactly once to allow detection of consonant errors, which were critical to our dependent measure.

2.3.1.2.1 *Sound Inventory.* Half of the tongue-twisters that each participant saw contained the /ɪ/ vowel (spelled ‘i’), and half contained the /æ/ vowel (spelled ‘a’). Given

[3] Due to changes in statistical model structure in the interim, as well as unanticipated exclusions, this power analysis may not reflect the actual power of the experiment reported herein to detect a significant effect.

[4] Although most syllables in this study were non-words, there were some real words among them due to the small number of unique syllables (127) that could be generated using the limited sound inventory and associated restrictions. We did not expect lexical effects to influence results, because the specific syllables used in each condition were counterbalanced across participants; only 36 (28.35%) of our syllables have meanings listed in WordNet (Feinerer & Hornik, 2020; Fellbaum, 1998; Wallace, 2007); and prior evidence suggests that the lexical bias effect is suppressed in a nonword context (Hartsuiker et al., 2005). Moreover, Goldrick (2004) found that results in single-day (first-order) phonotactic learning are robust to removal of word outcomes. In order to maximize power, we retained all word outcomes in our analysis of these experiments.

English's semi-opaque mapping between orthography and vowel pronunciation, we gave every participant two practice trials at the beginning of each participant's first session to clarify pronunciation of these vowels. These practice trials each contained one of these vowels and none of the target consonants in the experiment.

The consonant inventory consisted of (a) *unrestricted consonants* /m, n, k, g/, which could appear in any syllable position; (b) *English-restricted consonants*, which were restricted to particular syllable positions by virtue of English phonotactics (/h/ restricted to onset, /ŋ/ [spelled 'ng'] restricted to coda); (c) *experimentally-restricted consonants* /s, f/, which were the targets of the phonotactic generalizations under study here, and were restricted according to each participant's experimental condition and counterbalancing condition.

All sounds in the inventory whose International Phonetic Alphabet (IPA) symbols were part of the English language were written with their respective phonetic symbols, with the following exceptions:

- /g/ was written as 'gh' before 'i,' because in English, the letter 'g' can also be pronounced as /dʒ/ in this position (as in "ginger")
- /s/ was written as 'ss' at the end of syllables, because in English, the letter 's' can also be pronounced as /z/ in this position (as in "has" and "his")

2.3.1.2.2 ***Phonotactic Dependencies.*** Participants assigned to the *language-similar* condition saw stimuli in which experimentally-restricted consonants were restricted to *coda* when they appeared in the same syllable as a certain constraining vowel, which differed by consonant. In contrast, participants assigned to the *language-dissimilar* condition saw stimuli in which experimentally-restricted consonants were restricted to *onset* when they appeared in the same syllable as their respective constraining vowel. We randomly counterbalanced the mapping between constraining vowels and targeted consonants across participants, orthogonal to their experimental condition. Table 1 shows examples of stimuli in all conditions.

Table 1

Examples of Experiment 1 Tongue-Twister Stimuli

Experimental condition	Counter-balancing condition	Stimulus vowel	Example tongue-twister	Illegal sub-sequence
language-similar (vowel-coda restrictions)	i→s; a→f	i	**mis** hig fin king ● restricted-constrained: /s/ ● restricted-free: /f/	*si
		a	**maf** hag san kang ● restricted-constrained: /f/ ● restricted-free: /s/	*fa
	i→f; a→s	i	**mif** hig sin king ● restricted-constrained: /f/ ● restricted-free: /s/	*fi
		a	**mass** hag fan kang ● restricted-constrained: /s/ ● restricted-free: /f/	*sa
language-dissimilar (vowel-onset restrictions)	i→s; a→f	i	**sim** hig fin king ● restricted-constrained: /s/ ● restricted-free: /f/	*is
		a	**fam** hag san kang ● restricted-constrained: /f/ ● restricted-free: /s/	*af
	i→f; a→s	i	**fim** hig sin king ● restricted-constrained: /f/ ● restricted-free: /s/	*if
		a	**sam** hag fan kang ● restricted-constrained: /s/ ● restricted-free: /f/	*ass

Note. In all stimuli, unrestricted consonants were /m/, /n/, /g/, and /k/. Bulleted lists identifying restricted consonants are included for illustrative purposes only and were not shown to participants. Likewise, syllables demonstrating each phonotactic generalization are bolded and appear first here for clarity, but in the experiment itself, these syllables could appear anywhere in the tongue-twister and were not emphasized visually.

To construct stimulus lists for each condition, we first generated all 11,520 possible tongue-twisters in each condition. After removing tongue-twisters containing taboo syllables (‘fag’, ‘nig’), we randomly sampled 192 tongue-twisters to keep our sessions at a reasonable length of 96 trials each. Since the phonotactic generalization learned by each participant was the same on both days, we randomly sampled 96 tongue-twisters to present on Day 1 for each participant, and presented the remaining 96 tongue-twisters on Day 2.

2.3.1.3 Procedure. Upon submitting the online screening survey, participants were randomly assigned to either the *language-similar condition* (probability = $\frac{1}{3}$) or the *language-dissimilar condition* (probability = $\frac{2}{3}$) and informed how many days they would be asked to participate (two for the language-similar condition; four in the language-dissimilar condition). No other details about their participant assignment were revealed. Participants were then emailed a link to the consent form for the study. After providing their informed consent, participants booked appointments to pick up and return the study equipment (see “Sleep Data Collection,” Section 2.3.1.3.3 for description) in person at the Northwestern University campus and optionally provided their height, weight, age, and gender (for use in Fitbit’s sleep staging algorithm). Pick-up and drop-off appointments were scheduled at least five days apart, in order to allow participants time to complete all experimental sessions, including those in Experiment 2, between appointments. At the pick-up appointment, an experimenter who was unaware of each participant’s condition distributed the study equipment, told them to begin each experimental session 24 hours apart, and instructed them to wear the activity monitor while sleeping between sessions. The experimenter sent each participant a link to the first session of the study no later than 30 minutes after the pick-up appointment. All experimental sessions were completed online.

Due to the COVID-19 pandemic, we used participants' home computers and microphones to display stimuli and record responses, using a custom-built web application hosted on a Google Firebase server.[5] Due to data quality issues detected with Bluetooth microphones during piloting, all participants were instructed to use a non-Bluetooth microphone (either wired, USB, or built-in) to record their responses during experimental sessions. Initially, we instructed participants to use the Firefox, Chrome, Microsoft Edge, or Safari browser to access the study, but due to persistent audio quality issues with speech data from participants who used the Safari browser (see "Audio Data," Section 2.3.2.3.1), we instructed participants after January 19, 2022 ($N = 17$) to use only Firefox, Chrome, or Microsoft Edge.

Each experimental session lasted approximately 30 minutes, and sessions were gated such that participants could begin Day 2 no sooner than 24 hours after having begun Day 1. Each session began with an interactive wizard to verify that the participants' microphone was set up to record at the correct volume. Participants were then asked to read tongue-twisters to the beat of a metronome. In order to familiarize participants with the instructions and trial structure, each participant experienced two practice tongue-twister trials after the microphone set-up phase at the start of Day 1, and one at the start of Day 2 as a reminder. These practice trials did not contain any of the critical consonants from the rest of the experiment, and were identical across conditions. Each session contained 96 trials in random order (1 tongue-twister per trial).

2.3.1.3.1 *Trial Structure.* At the start of each trial, a fixation cross appeared on the screen, accompanied by an initial low-pitched (700 Hz) beep to signal the beginning of the trial, and the participant's microphone began recording. One second later, the tongue-twister stimulus appeared written on the screen with spaces separating each syllable, accompanied by a

[5] Code for this application is available at https://github.com/nimirea/tongue-twisters-online-firebase.

medium-pitched (882 Hz) “get ready” beep. One second after the appearance of the stimulus, the metronome began to play at a rate of 1 beat per second, set to a higher pitch (1049 Hz) to distinguish it from the “get ready” beep. After 4 metronome beats (1 repetition of the tongue-twister, the metronome sped up to a rate of 2.53 beats per second, and participants were instructed to repeat the tongue-twister three times at this faster rate. After the 12 fast beats, the microphone stopped recording, a clickable “next” button appeared beneath the tongue-twister (enabling participants to take self-timed breaks before continuing to the next trial), and the individual trial recording was uploaded to the server in Waveform Audio File Format.

2.3.1.3.2 *Surveys.* At the end of each experimental session, participants were asked what equipment they used for the session (browser, computer operating system, headphones, and microphone) to help diagnose data quality issues during data collection. They were also asked how many hours of sleep they received the previous night, and had an opportunity to provide other written feedback about their experience. The Karolinska Sleep Questionnaire (KSQ; Westerlund et al., 2014) was additionally administered after Day 1’s session.

2.3.1.3.3 *Sleep Data Collection*. We collected sleep data using Fitbit Inspire 2 wrist-worn activity monitors, which use body movement, heart rate, and heart rate variability to detect sleep/wake intervals and identify different sleep stages (Stage 1, Stage 2, and Deep [Slow-Wave] Sleep). When measured against an EEG machine, Fitbit’s algorithms using these variables are comparable to a standard wrist-worn actigraphy device in terms of identifying slow-wave sleep and detecting sleep intervals (Haghayegh et al., 2020). Participants were asked to wear these activity monitors on nights between sessions, on their non-dominant hand, one finger-width up the wrist, as recommended by Fitbit’s support website (Fitbit, Inc., n.d.).

2.3.2 *Results*

2.3.2.1 Data Processing. Two transcribers independently transcribed recordings from this experiment as well as from Experiment 2, with parallel coding of 15% of trials for reliability. To ensure that transcribers were naive to experimental condition, each session was given its own randomly generated identification number before transcription, and the mapping between random session ID, participant, and day was stored in a separate file that neither transcriber accessed until after transcription had been completed. To diagnose microphone quality issues before transcription, five randomly-selected trials from all sessions were checked by the first author; sessions were excluded if less than 12 syllables in each of these trials were clearly audible (17 sessions, resulting in 11 excluded participants).

2.3.2.2 Statistical Data Analysis. To test the hypothesis that prior knowledge provides a consolidation benefit, we built a mixed-effects logistic regression that predicted position-preservation of consonants produced in participants' errors from session, condition, type of consonant produced, and context of consonant produced. We used the lme4 package (Bates, Mächler, et al., 2015) in the R programming language (R Core Team, 2022) to fit this model.

2.3.2.2.1 ***Dependent Measure.*** Our learning measure of interest was the rate of syllable position maintenance during "fast" repetitions of each tongue-twister, and so the dependent measure in our statistical model was whether each error preserved the syllable position of the consonant that was produced in the error, relative to where it appeared in the stimulus tongue-twister (1) or not (0).

We used the following automated procedure to extract and classify errors from transcriptions:

1. We split the transcription of each produced syllable into onset, vowel, and coda.

2. We truncated consonant clusters to include only the first consonant of the cluster.
3. Each produced syllable was aligned to one of the four syllables in the target stimulus via a similarity metric that took shared onsets and codas into account equally, as well as a fallback heuristic involving syllable position. For each syllable produced in a repetition, we calculated the total number of onset and coda consonants it matched with each syllable in the target tongue-twister. If a single target syllable had more matches than all other target syllables, we chose that target syllable as the produced syllable's aligned target. Otherwise, the produced syllable was aligned with the target syllable in the closest position; if two target syllables were equally distant, we selected the first target syllable in the tie set as the aligned target.
4. We identified all consonants that did not match their aligned targets as errors.
5. We classified each error as either position-preserving if the position of the produced consonant matched the consonant's position in the target stimulus, or as position-changing otherwise. Out-of-vocabulary consonants and other sounds ($N = 464$) were removed from the dataset, since they could not be classified as either position-preserving or position-changing.

Based on an inspection of the data, it was clear that coda targets were more likely to preserve position than onset targets. Therefore, to create a fair baseline within each condition, we only analyzed errors made in positions to which the experimentally-restricted consonants were constrained—*e.g.*, for participants that learned phonotactics restricting items in coda (language-similar condition), we only included errors in our model where produced consonants were in coda position in the target, and we only included errors where produced consonants were in onset position for participants in the language-dissimilar condition.

2.3.2.2.2 *Fixed Effects.* Fixed effects were session, condition, and two factors encoding contrasts among consonant types in context, and interactions between all factors except the factors encoding contrasts among consonant types in context.

Contrast coding for each factor was centered on zero, such that the intercept of the model reflected the grand mean. We defined *session* as a within-participant factor with two effect-coded levels: pre-consolidation (Day 1), coded as -0.5, and post-consolidation (Day 2), coded as 0.5. We defined *similarity to English* as a between-participants factor with two effect-coded levels: similar, coded as 0.5, and dissimilar, coded as -0.5.

In contrast to previous second-order phonotactic learning studies that constrained experimentally-restricted consonants to a certain position in every trial, our study only constrained experimentally-restricted consonants to a certain position in the presence of their respective constraining vowel. Our statistical model therefore took into account differences in context among experimentally-restricted consonants by using a two-factor Helmert coding scheme (see Table 2), with separate fixed effects for the difference between experimentally-restricted consonants and unrestricted consonants (*unrestricted vs. restricted*) and the difference between experimentally-restricted consonants in whose position was constrained in the current vowel context and experimentally-restricted consonants whose position was not restricted in the current trial (*restricted-constrained vs. restricted-free*).

Table 2

Contrast Coding Scheme for Consonant Type and Context

Consonant type	Constraining context	Number of consonants per trial	Fixed effect coding	
			unrestricted vs. restricted	*restricted-constrained vs. restricted-free*
experimentally-restricted	constrained = in the context of its conditioning vowel	1	-0.66	0.5
	unconstrained = in the context of the other (non-conditioning) vowel	1	-0.66	-0.5
unrestricted	N/A	4	0.33	0

Note. Predictors used in the model to encode consonant type and context. Each fixed effect was coded such that the sum of values of across consonants was zero.

2.3.2.2.3 *Random Effects Structure, Convergence, and Parsimony.* In addition to aforementioned fixed effects, we initially included random intercepts and slopes for all fixed effects by participant.[6] We used the BOBYQA optimizer (Powell, 2009) to fit model parameters and perform convergence checks.

Because the maximal model did not converge within 100,000 iterations, we constrained correlation parameters to zero, following the recommendations in Bates et al. (2015) to avoid

[6] In our pre-registration, we initially stated that we would use the maximal random effects structure. However, due to the relatively homogenous pool from which stimulus items were chosen, as well as the comparatively small number of observations per tongue-twister (*M* = 7.66, *SD* = 5.21), relative to the number of observations per participant (*M* = 77.72, *SD* = 66.13), we included only by-participant random effects in the model, opting to forgo by-stimulus random effects.

overparameterization.[7] To avoid singular fit, we removed all third-order random slopes, as well as the random by-participant slope of the interaction between *restricted-constrained vs. restricted-free* and *session*.

2.3.2.2.4 ***Significance Testing.*** The significance of each predictor was assessed via a likelihood ratio test, in which we compared the likelihood of the data under the full model to the likelihood of the data under a model that did not include the predictor in question via a χ^2 test.

2.3.2.3 Data Quality.

2.3.2.3.1 ***Audio Data.*** A randomly selected set of 16 sessions (1532 trials) was transcribed by both transcribers. Overall, transcribers agreed on 96.63% ($N = 35{,}684$) of all consonants transcribed (97.39% of onsets and 94.87% of codas) in this set, with Cohen's κ = 0.97 on all consonants, indicating almost perfect agreement (Landis & Koch, 1977). The most frequently confused pair of consonants was /m/ and /n/, with 111 occurrences of disagreement between transcribers.

When looking only at errors, however, transcribers only agreed on 62.15% ($N = 1{,}000$) of consonants, with Cohen's $\kappa = 0.57$, indicating only moderate agreement (Landis & Koch, 1977); agreement on onset errors was 62.58% (Cohen's $\kappa = 0.56$), while agreement on coda errors was 61.97% (Cohen's $\kappa = 0.55$). These discrepancies in agreement between transcriptions of errors and transcriptions of all consonants may be due to (a) the smaller number of observations (36,928 total vs. 2,622 in errors); (b) the fact that virtually all disagreements regarding the identity of consonants (97.49%; $N = 824$) involve substituting either an errorful consonant for a correct one, or the exchange of two errorful consonants; and (c) the fact that more nearly a third

[7] We followed the recommendations in (Bates, Kliegl, et al., 2015) with a single exception: we did not further reduce the random effects structure beyond the number of principal components that cumulatively accounted for 100% of the variance in random effects. This is the most conservative version of the model.

(29.23%; $N = 178$) of disagreements on errorful consonants were due to disagreements regarding splitting up a trial into repetitions—that is, these consonants were transcribed by both transcribers, in the same syllable position, but were not assigned to the same repetition and/or location within a repetition.

Each transcriber rated audio quality for each transcribed session from 1 (worst) to 5 (best); the most common audio quality rating was 5 ($N = 56$ sessions). We excluded 11 participants from transcription due to audio quality issues.

2.3.2.3.2 *Sleep Data.* We collected biometric data from 71 participants between Day 1 and Day 2 using activity trackers. However, of these, only 51 participants (71.83%) provided their height, weight, age, and gender[8] when signing up for the study. Since Fitbit uses this demographic information to calculate heart rate zones for each participant, (M. Considine, personal communication, March 19, 2021), we omitted the 20 participants who did not provide full demographic information from our analysis of sleep data.

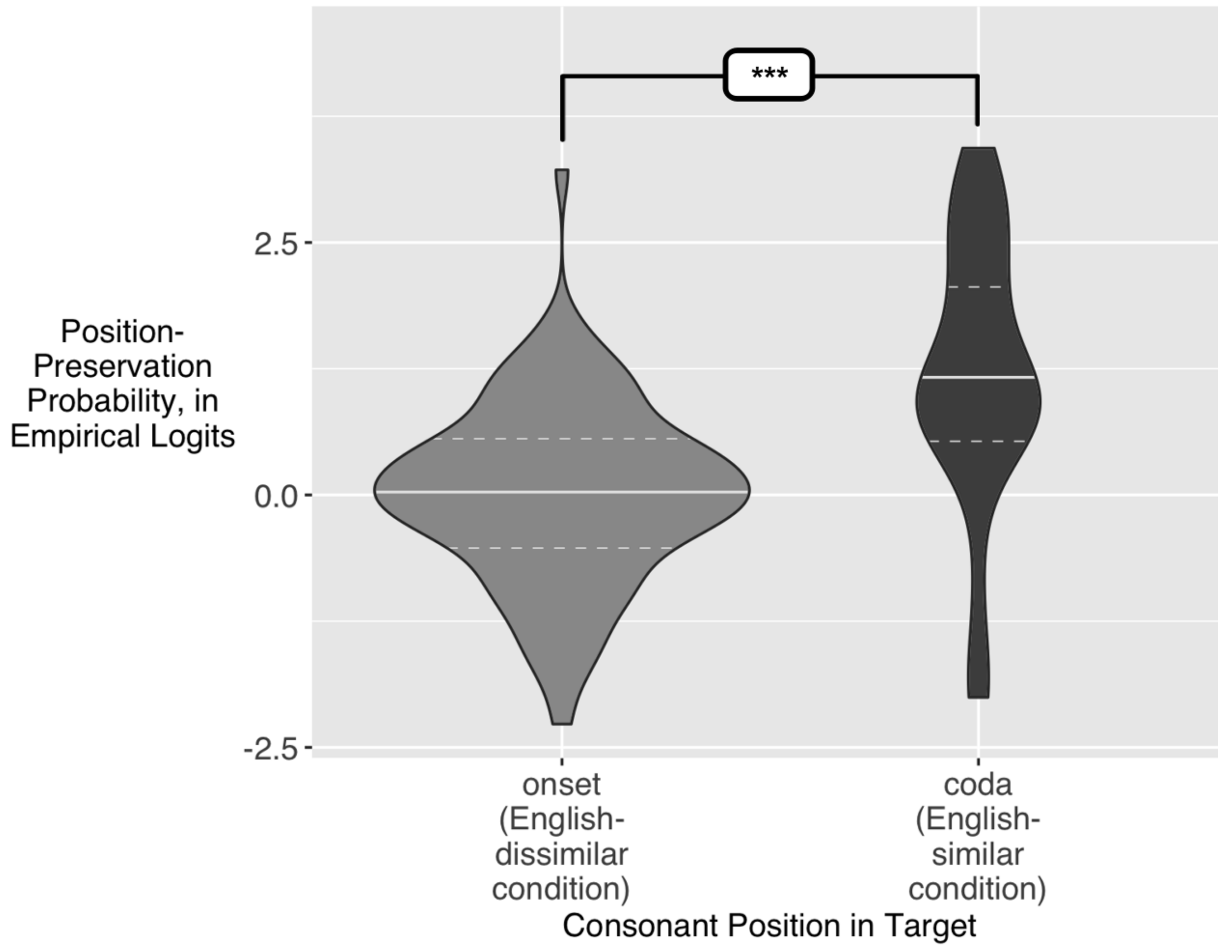
2.3.2.4 Speech Error Position Analysis. On average, approximately 53.40% of consonants produced in errors preserved their syllable positions ($\beta = 0.73$, SE = 0.12). We found significant main effects of *similarity to English*, *unrestricted vs. restricted*, and *restricted-constrained vs. restricted-free* (see Figures 1 and 2 for visualizations of these effects), with models including all predictors except each of these resulting in significant reductions in fit to the data compared to the full model, according to chi-squared tests (see Table 3 for

[8] We recorded gender as a free-response variable in our set-up survey, but the Fitbit sign-up form only had two possible choices for gender: *Male* and *Female*. We coded responses as *Female* if participant responses to the gender question were “F” or “f”, or included “female”, “Female”, or “woman” ($N = 38$); *Male* if participants responded “M”, “m”, or if their response contained “Male”, “male”, or “man” ($N = 11$); participants who indicated another gender were assigned to one of the two binary genders with equal probability using a random number generator ($N = 2$).

chi-squared test results and coefficient estimates from the statistical model). When visualizing the results of this and subsequent experiments, we transformed position-preservation rate using the empirical logit function described in Appendix A.

Figure 2

Probability of Errors in Experiment 1 Preserving Position, by Condition

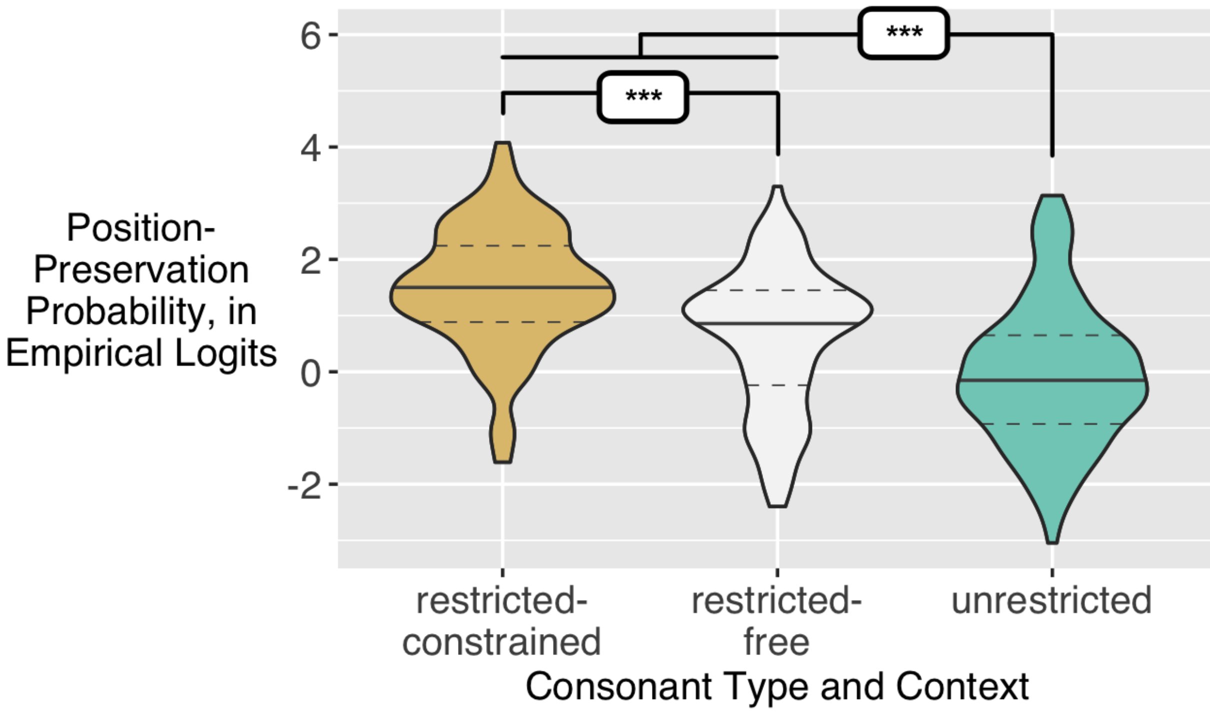


Note. Position-preservation probability was calculated for experimentally-restricted and unrestricted consonants in each session; area of each violin represents the total number of sessions in each condition in which at least one error on both experimentally-restricted and unrestricted consonants was made. Horizontal lines show quantiles. Significance marks are derived via the significance tests described in Section 2.3.2.2.4.

***$p < 0.001$

Figure 3

Consonants' Probability of Preserving Position in Errors Made in Experiment 1



Note. Position-preservation probability was calculated for each session; area of each violin represents the total number of sessions in which at least one error involving a consonant of the respective type-context combination was made. Horizontal lines show quantiles. Significance marks are derived via the significance tests described in Section 2.3.2.2.4.

***$p < 0.001$

Table 3

Experiment 1 Speech Error Mixed-Effects Model Results

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	0.73	0.12		
session	0.01	0.13	0.01	0.918
restricted-constrained vs. restricted-free	1.08	0.28	13.76	< 0.001 ***
unrestricted vs. restricted	-1.83	0.18	66.94	< 0.001 ***
similarity to English	1.40	0.25	27.99	< 0.001 ***
restricted-constrained vs. restricted-free × session	-0.29	0.42	0.46	0.498
restricted-constrained vs. restricted-free × similarity to English	-0.39	0.55	0.43	0.513
unrestricted vs. restricted × session	-0.21	0.3	0.41	0.521
unrestricted vs. restricted × similarity to English	-0.33	0.35	0.89	0.345
similarity to English × session	-0.14	0.27	0.27	0.606
restricted-constrained vs. restricted-free × similarity to English × session	-0.81	0.84	0.9	0.343
unrestricted vs. restricted × similarity to English × session	0.14	0.6	0.05	0.817

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4.

All else being equal, the odds of errors from participants in the similar-to-English condition preserving position were approximately 4.06 (β = 1.40, SE = 0.25) times those of errors made by participants in the dissimilar condition preserving position (see Figure 2). Since this applied to unrestricted as well as restricted consonants, we interpret this as an effect of syllable position, rather than an effect driven by training condition (see discussion before Figure 4).

The odds of unrestricted consonants preserving position in errors were approximately 0.16 (β = -1.83, SE = 0.18) times those of restricted consonants preserving position in errors, which in turn had odds of preserving position that were approximately 2.96 (β = 1.08, SE = 0.28) higher when they appeared in their constrained context compared to their free context (see Figure 3), indicating that participants successfully learned the second-order phonotactic restrictions presented during the experiment.

In order to test whether there was a significant difference in position-preservation rates between restricted consonants in free contexts and unrestricted consonants, we re-analyzed our data using a version of the model that set restricted consonants in free contexts as the baseline in a simple coding scheme (see Appendix B for details of model and results). This analysis revealed a significant difference in the odds of position-preservation between unrestricted and restricted consonants in free vowel contexts, with the odds of unrestricted consonants preserving position being 0.27 times those of restricted consonants in free vowel contexts preserving position (β = -1.27, SE = 0.21; $\chi^2(1) = 32.53$, $p < 0.001$). This indicates that participants generalized the restriction to unconstrained contexts.

We failed to find any significant interactions between predictors, or any significant main effect of *session*. In a follow-up analysis on Day 1 data only (see Appendix D for details) we detected a significant effect of both the *unrestricted vs. restricted* (β = -1.53, SE = 0.26; $\chi^2(1)$ = 23.42, $p < 0.001$) predictor and the *restricted-constrained vs. restricted-free* predictor (β = 1.19, SE = 0.40; $\chi^2(1)$ = 7.84, $p < 0.01$), indicating that participants were indeed using the phonotactic restrictions to guide their productions as early as Day 1.

We failed to replicate Westerlund et al. (2013)'s correlation between slow-wave sleep and self-reported restoration for sleep (see Appendix C for a more thorough treatment of this data).

2.3.2.4 Number of Errors. On average, participants produced 100.84 errors per session (minimum = 9; maximum = 328). In order to test whether the number of errors in each session may have mediated the effects we found in our analysis, we first tested whether the number of errors per session correlated with either *similarity to English* or *session* by fitting a negative binomial linear model that predicted number of errors from *similarity to English*, *session*, and their interaction. Using the model simplification procedures described in Bates et al. (2015), we were able to retain random intercepts by participant in the model. We used the same significance testing procedure described in Section 2.3.2.2.4 to test whether each predictor was significant. Results are shown in Table 4.

Table 4

Experiment 1 Number of Errors Model Results

Predictor	Estimate	Standard Error	$\chi^2(1)$	*p* value of χ^2 test
(Intercept)	4.35	0.10		
session	-0.20	0.05	15.30	< 0.001 ***
similarity to English	-0.05	0.19	0.07	0.781
similarity to English × session	-0.03	0.10	0.10	0.757

Note. The $\chi^2(1)$ and *p* values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4.

We found significant correlations between *session* and the number of errors in each session. Participants made significantly more errors on Day 1 (M = 108.31, SD = 74.24) than on Day 2 (M = 93.37, SD = 73.36). Since we did not find any significant effects of *session* or interactions between *session* and any other variables in our main model, we conclude that the results of our main model were not caused by differences in the number of errors across participants.

2.3.3 *Discussion*

In this experiment, we tested whether the benefits of consolidation are modulated by similarity to participants' native language, in order to assess whether consolidation serves to integrate newly learned phonotactic knowledge with phonotactic knowledge from participants' native language. We did not find evidence of this differential benefit. However, we did find evidence of phonotactic learning in this experiment, with restricted consonants preserving position more often than unrestricted consonants in errors, and restricted consonants preserving position more often when they appeared with their constraining vowels than in their unconstrained contexts. Errors made by participants learning a vowel-coda restriction also preserved position more often than those made by participants who learned a vowel-onset restriction, though we believe this is ultimately due to the tendency of English speakers to preserve rhyme structure in errors.

We found that restricted consonants preserved syllable position more often in errors than unrestricted consonants—a result that has been interpreted as phonotactic learning in previous research. Additionally, restricted consonants were significantly more likely to preserve position when they appeared in their constrained context, meaning that participants were able to successfully bind consonants to their respective constraining vowel contexts. Taken together, these results indicate that participants did use the statistical regularities present in the stimuli to guide their productions. We take this as evidence of successful learning of second-order phonotactic restrictions. However, in our follow-up analysis (see Appendix B), we did discover that restricted consonants preserved position significantly more often than unrestricted consonants, even in vowel contexts where the restricted consonants' positions were unconstrained. If participants' errors perfectly reflected the statistical characteristics of the stimuli, errors on restricted consonants in unconstrained vowel contexts should preserve position just as often as unrestricted consonants. Instead, the higher position-preservation rates for restricted consonants, even in unconstrained vowel contexts, indicate that participants generalize positional restrictions across vowel contexts, for a given segment. Participants' errors therefore reflect overlapping influence of global, first-order phonotactic generalizations (e.g., "/f/ appears in coda position 75% of the time, throughout the experiment") and local, second-order phonotactic generalizations (e.g., "/f/ is more likely to appear in coda position, if /ɪ/ is the vowel in this syllable")—suggesting that participants may be learning both types of statistical generalizations simultaneously.

Despite our considerable sample size (*N* = 68 participants total), we did not detect any significant interactions between session and either restrictedness or constraining vowel context. In fact a follow-up analysis on Day 1 data only (see Appendix D for details) revealed that participants were making use of phonotactic restrictions on Day 1. This was unexpected given the results of several past experiments showing that second-order phonotactics only begin to affect undistracted adults' productions after a period of sleep (Gaskell et al., 2014; Smalle et al., 2017, 2021; Warker, 2013; Warker et al., 2010; Warker & Dell, 2006). The discrepancy between our results and those of past studies suggests that this experiment's design may have introduced a key difference that enabled participants to make use of the phonotactic restrictions on Day 1. One possibility is that the remote modality of this study may have led to higher levels of participant distraction, and that these distraction may have introduced higher cognitive load, such that participants in our study resembled the cognitively-depleted condition from Smalle et al. (2021). This is unlikely, however, since we excluded all participants who took longer than one hour to complete either session, and only 16.67% of participants (*N* = 7) took more than the estimated 30 minutes to complete Day 1.

Another possibility is that the unique characteristics of our stimuli may have made them easier to utilize in production on Day 1. Participants in past second-order phonotactic learning studies have learned sets of phonotactic restrictions that constrain two consonants to "opposite" syllable positions, depending on the identity of the vowel—for example, "If /ɪ/ is the vowel, then /s/ can only be an onset, and /f/ can only be a coda; if /æ/ is the vowel, then /s/ can only be a coda, and /f/ can only be an onset." In this study, by contrast, each participant learns a set of restrictions in which vowels only restrict a single consonant—for example, in the *similar-to-English* condition: "If /ɪ/ is the vowel, then /s/ can only be an coda (and /f/ can be either onset or coda); if /æ/ is the vowel, then /f/ can only be a coda (and /s/ can be either onset or coda)." In this example, syllables such as "CVs" and "CVf" are permitted no matter what vowel is chosen for "V," which may remove the incompatibility that hidden units serve to overcome. In other words, participants learn that /s/ and /f/ are permitted in coda regardless of the vowel, allowing the production system to "ignore" the identity of the vowel and treat this restriction like a first-order phonotactic restriction that constrains both /s/ and /f/ to syllable coda.

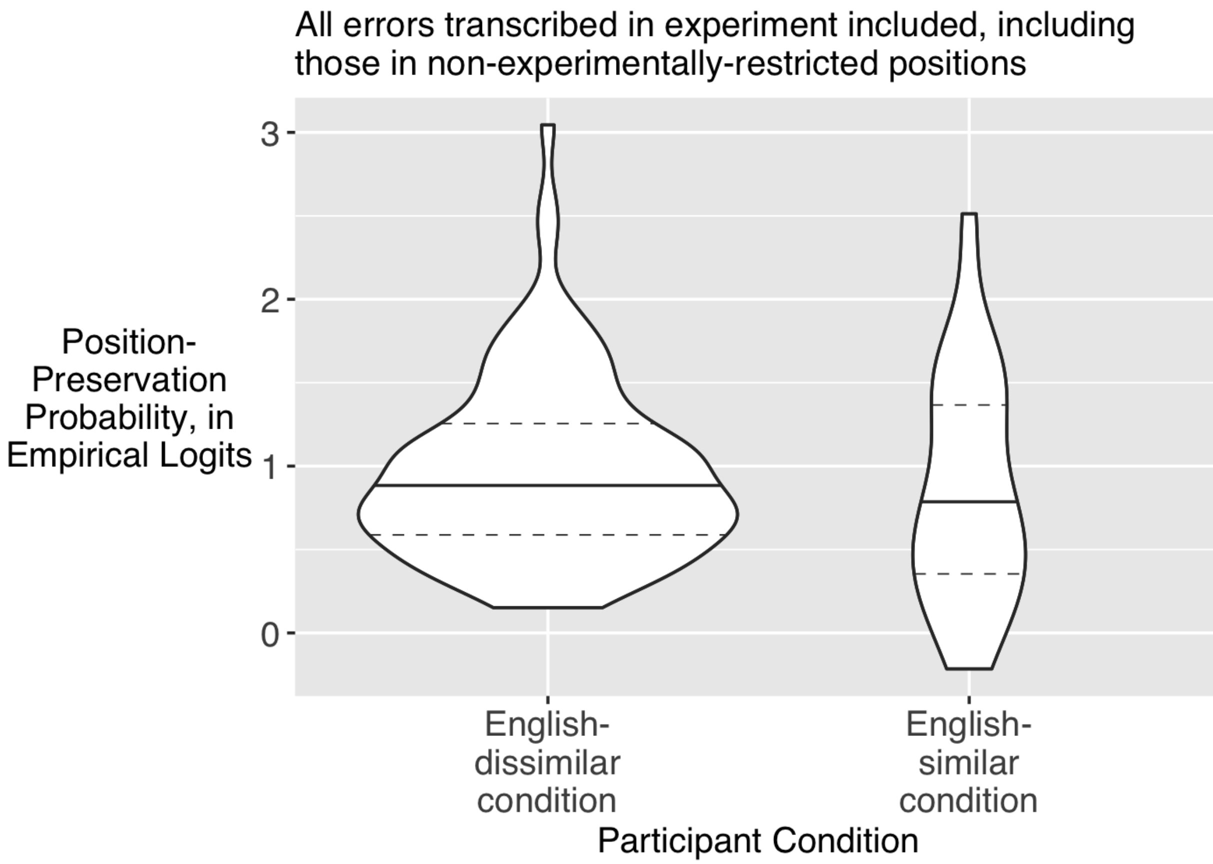
Nevertheless, hidden units (at least, as formulated by Dell et al., 2021 and Warker & Dell, 2006) are still required to represent the difference among experimentally-restricted consonants in constrained vs. unconstrained vowel contexts, so this account does not fully explain the lack of a consolidation effect for the difference among vowel contexts. It is possible that the design of our study allowed us to gather comparatively less evidence for each phonotactic restriction, since each experimentally-restricted consonant was only constrained to a specific position in 50% of trials (48 for each session). This interaction may yet emerge if we were to increase the power of the study by doubling the number of trials per day, such that each experimentally-restricted consonant would be constrained in 96 trials per session.

We also found a significant main effect of *similarity to English*, such that errors made by participants in the similar condition preserved position significantly more often than those made by participants in the dissimilar condition. Recall that we only analyzed errors made on consonants that originally appeared in the constrained position (coda for the *similar-to-English* condition, onset for the *dissimilar-to-English* condition). Therefore, all position-preserving errors analyzed in the *similar-to-English* condition were errors in which the rhyme of the target syllable remained intact. The higher rate of position-preservation in this condition, then, may reflect a general tendency of English speakers to preserve the structure of the rhyme in errors. This tendency is present in errors on a short-term memory task (Lee & Goldrick, 2008) and weakly attested in previous work using the tongue-twister paradigm (Dell et al., 2000; Laubstein, 1987; though see Goldrick & Larson, 2008, for an exception involving /s/). Further investigation is necessary to determine the scope of this tendency. In our data, however, we have two related pieces of evidence in this direction:

1. When unanalyzed errors (i.e., those occurring in the unconstrained position) are plotted together with the analyzed errors, the difference between conditions disappears (see Figure 4). This suggests that the effect is driven mostly by our choice to analyze codas in the English-similar condition, and onsets in the English-dissimilar condition.
2. For unrestricted and language-restricted consonants, whose positions were not affected by our experimental manipulations, errors on coda targets generally preserved position more often than errors on onset targets, in both conditions (see Figure 5). This main effect in our statistical analysis reflected this difference, since we compared onset consonants in the English-dissimilar condition to coda consonants in the English-similar condition, even though these consonants appeared in the same positions with the same frequency in both conditions. This led us to conclude that the main effect of condition was due to a more general property of English speakers to preserve coda position more often than onset position in errors, across all target consonants.

Figure 4

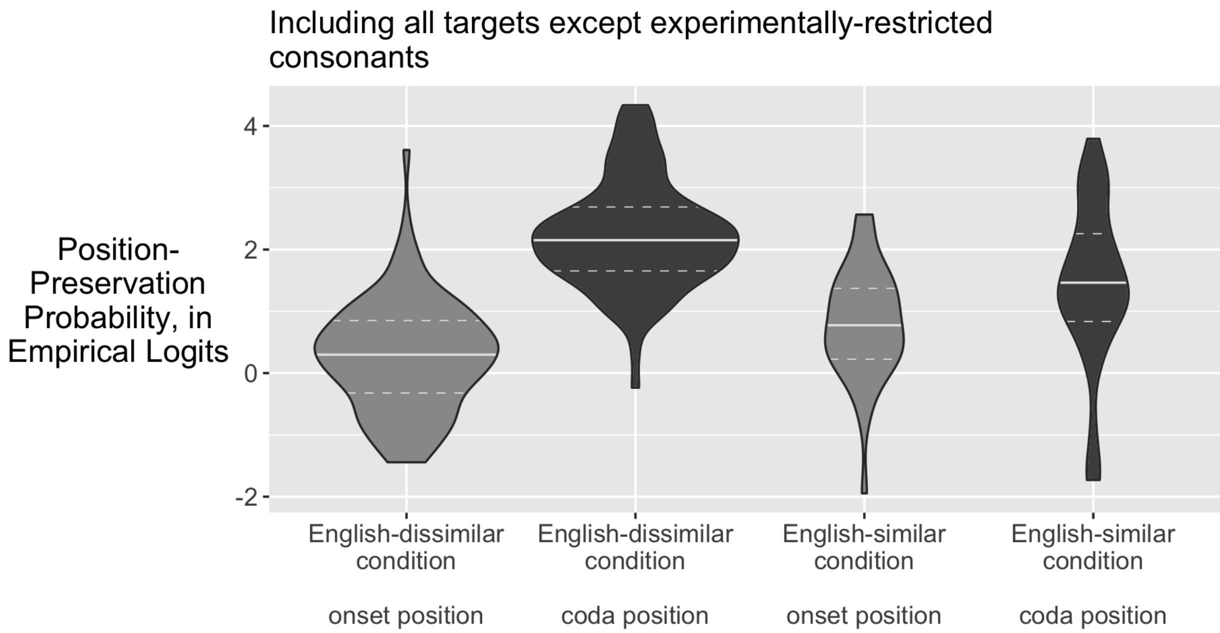
Position-Preservation Rates in Experiment 1, Including All Transcribed Consonants



Note. Area of each violin represents the total number of sessions in each condition. Horizontal lines show quantiles.

Figure 5

Coda Position-Preservation Bias across Non-Experimentally-Restricted Consonants in Experiment 1



Note. Area of each violin represents the total number of sessions in each condition in which at least one error on a consonant in the respective position was made. Horizontal lines show quantiles. Only categories on the outside of the figure (English-dissimilar + onset, English-similar + coda) were included in the statistical analysis.

2.4 Experiment 2: Prior Knowledge from Pre-Training

2.4.1 *Method*

2.4.1.1 Participants. We randomly selected 79 participants to participate in this experiment after taking part in the *language-dissimilar* condition of Experiment 1, which served as the pre-training phase for this experiment. Compensation for this experiment alone was $24 ($11 for Day 1 + $13 for Day 2).

A total of 39 participants were excluded due to poor audio quality ($N = 12$), attrition during the pre-training phase ($N = 8$), attrition between Days 1 and 2 ($N = 3$), experiment software malfunction ($N = 9$), failing to sleep for at least 3 hours between any sessions ($N = 4$), not making at least one error on the relevant syllable positions for analysis (see "Dependent Measure," Section 2.3.2.2.1) in both sessions ($N = 4$), taking more than one hour to complete any session ($N = 3$), experimenter error in verifying eligibility requirements ($N = 3$), and/or audible distractions during the pre-training phase ($N = 2$), for a final sample size of 40 participants.

2.4.1.2 Materials. Stimuli were constructed as in Experiment 1 (see Section 2.3.1.2 for details), with two exceptions: the identity of experimentally-restricted consonants and the phonotactic dependencies learned.

2.4.1.2.1 *Sound Inventory*. All sounds used in pre-training were also used in this experiment, with the exception of the previously experimentally-restricted consonants /s/ and /f/. Instead, /z/ and /v/ were introduced as experimentally-restricted consonants, in order to distinguish stimuli from pre-training while creating an opportunity for transfer through similar phonetic features (see Table 5 for examples; Goldrick, 2004, for evidence of within-session transfer based on shared features).

Table 5

Examples of Experiment 2 Tongue-Twister Stimuli

Experimental condition	Counterbalancing condition	Stimulus Vowel	Example tongue-twister	Illegal subsequence
training-similar (vowel-onset restrictions)	i→z, a→v	i	**zim** hig vin king	*iz
		a	**vam** hag zan kang	*av
	i→v, a→z	i	**vim** hig zin king	*iv
		a	**zam** hag van kang	*az
training-dissimilar (vowel-coda restrictions)	i→z, a→v	i	**miz** hig vin king	*zi
		a	**mav** hag zan kang	*va
	i→v, a→z	i	**miv** hig zin king	*vi
		a	**maz** hag van kang	*za

Note. Examples of tongue-twisters in each condition of Experiment 2. Syllables demonstrating the phonotactic generalization in question are bolded and appear first here for clarity, but in the experiment itself, these syllables may appear anywhere in the tongue-twister and are not emphasized visually. Illegal subsequences never appeared in tongue-twisters shown to participants in the respective condition.

For each participant, the constraining vowel of every experimentally-restricted consonant matched the constraining vowel of its voiceless counterpart in pre-training. That is, all participants who saw stimuli following "i→s, a→f" restrictions in the pre-training phase saw stimuli here that followed "i→z, a→v" restrictions; similarly, participants who saw stimuli following "i→f, a→s" restrictions in the pre-training phase saw stimuli here that followed "i→v, a→z" restrictions.

2.4.1.2.2 *Phonotactic Dependencies.* At the start of the study, participants who took part were randomly assigned to either the *training-dissimilar* or the *training-similar* condition, with equal probability. During the two days of this experiment, participants in the *training-similar* condition saw stimuli in which experimentally-restricted consonants were restricted to onset when they appeared in the same stimuli as their respective constraining vowel, just as in the pre-training phase. Experimentally-restricted consonants in stimuli presented to participants in the *training-dissimilar* condition, on the other hand, were restricted to coda when they appeared in the same stimuli as their respective constraining vowel.

As in Experiment 1, in order to construct stimulus lists for each condition, we again generated all 11,520 possible tongue-twisters in each condition. After removing tongue-twisters containing taboo syllables ('vag', 'nig'), we randomly sampled 192 tongue-twisters. Since the phonotactic generalization learned by each participant was the same on both days, we randomly sampled 96 tongue-twisters to present on Day 1 for each participant, and presented the remaining 96 tongue-twisters on Day 2.

2.4.1.3 Procedure. All participants took part in this experiment at least 24 hours after beginning Day 2 of the pre-training phase (Experiment 1, language-dissimilar condition), to allow for a consolidation of the pre-trained vowel-onset restrictions. The procedure in this experiment was identical to that of Experiment 1 (see "Procedure," Section 2.3.1.3 for details), except that participants did not re-complete the KSQ on Day 3. Participants were randomly assigned to either the training-similar condition or the training-dissimilar condition with equal probability on Day 3.

2.4.2 *Results*

2.4.2.1 Data Processing. We transcribed speech data from this experiment simultaneously with that from Experiment 1, according to the method described in Section 2.3.2.1. We excluded data from participants who produced less than 12 clearly audible syllables in each of the five tester trials from any session (N = 27 sessions, resulting in 12 excluded participants).

2.4.2.2 Statistical Data Analysis. We used the same procedure for automating and classifying errors from transcriptions and statistical model as in Experiment 1 (see Section 2.3.2.2 for details), with the exception of the *similarity to English* factor, which we replaced with a *similarity to training* factor to reflect conditions in this experiment. This predictor was coded as -0.5 for errors from participants in the *training-dissimilar* condition, and 0.5 for errors from participants in the *training-similar* condition.

2.4.2.3 Audio Data Quality. A randomly selected set of 13 sessions (1248 trials) was transcribed by both transcribers. Overall, transcribers agreed on 96.35% (N = 28,904) of all consonants transcribed (96.99% of onsets and 95.71% of codas) in this set, with Cohen's κ = 0.96 on all consonants, indicating almost perfect agreement (Landis & Koch, 1977). The most frequently confused pair of consonants was /m/ and /n/, with 258 occurrences of disagreement between transcribers.

When looking only at errors, however, transcribers only agreed on 52.54% (N = 1014) of consonants, with Cohen's κ = 0.45 indicating only fair-to-moderate agreement (Landis & Koch, 1977); agreement on onset errors was 48.66% (Cohen's κ = 0.40), while agreement on coda errors was 54.98% (Cohen's κ = 0.47). These discrepancies in agreement between transcriptions of errors and transcriptions of all consonants were due to (a) the smaller number of observations

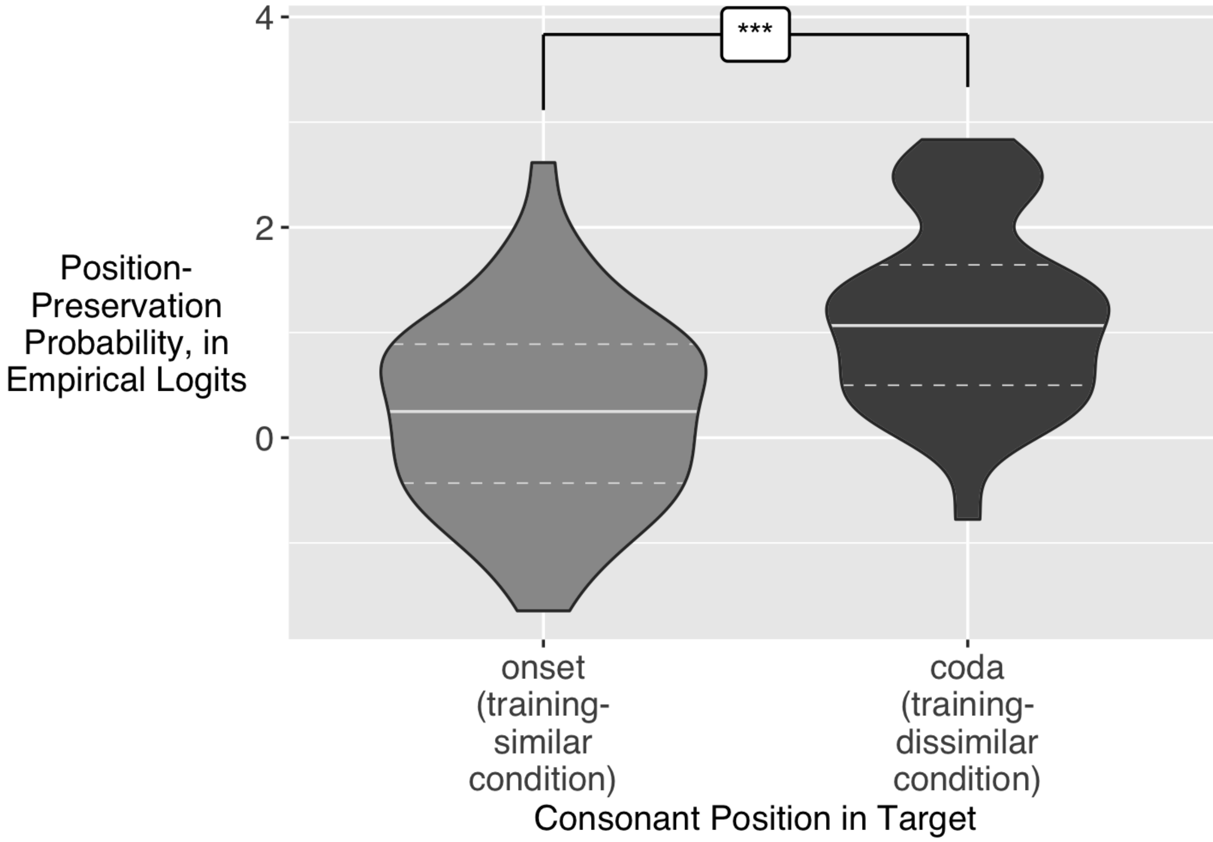
(29,998 total vs. 1,930 in errors); (b) the fact that almost all disagreements regarding the identity of consonants (93.23%; $N = 868$) involve substituting either an errorful consonant for a correct one, or the exchange of two errorful consonants; and (c) the fact that almost half (42.36%; $N = 388$) of disagreements on errorful consonants were due to disagreements regarding splitting up a trial into repetitions—that is, these consonants were transcribed by both transcribers, in the same syllable position, but were not assigned to the same repetition and/or location within a repetition.

Each transcriber rated audio quality for each transcribed session from 1 (worst) to 5 (best). Again, 5 was the most common audio quality rating.

2.4.2.4 Speech Error Position Analysis. On average, 69.42% of errors preserved their position across all participants and sessions ($\beta = 0.80$, SE = 0.13). We found significant main effects of *similarity to training* (Figure 6), as well as *unrestricted vs. restricted* and *restricted-constrained vs. restricted-free* (Figure 7), with models including all predictors except each of these resulting in significant reductions in fit to the data compared to the full model, according to chi-squared tests (see Table 6 for chi-squared test results and coefficient estimates from statistical model described in Section 2.4.2.2). All else being equal, the odds of errors from participants in the similar-to-training condition preserving position were approximately 0.38 ($\beta = -0.97$, SE = 0.26) times the odds of errors from participants in the dissimilar-to-training condition preserving syllable position (see Figure 6). This reflects the same tendency found in Experiment 1: codas (the consonants analyzed for participants in the *training-dissimilar* condition) are more likely to preserve position in errors than onsets (the consonants analyzed for participants in the *training-similar* condition).

Figure 6

Probability of Errors in Experiment 2 Preserving Position, by Condition

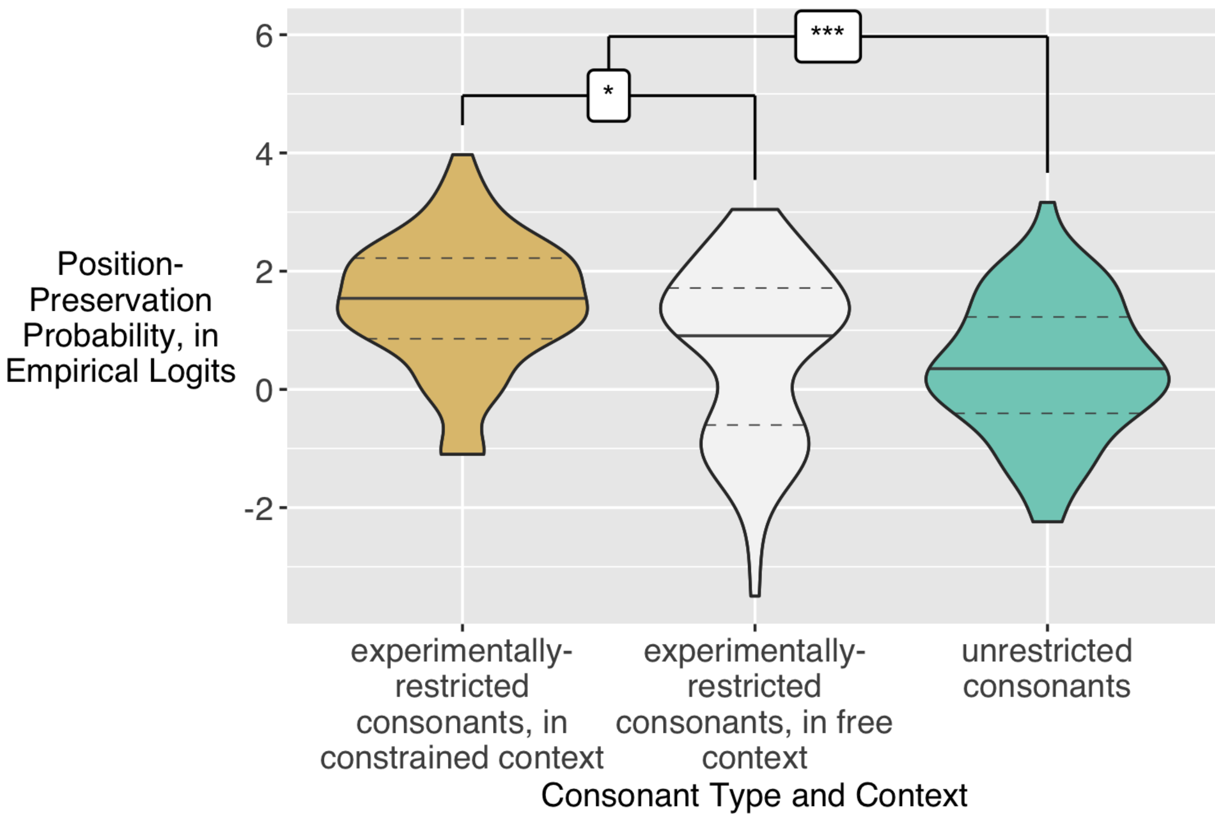


Note. Position-preservation probability was calculated for experimentally-restricted and unrestricted consonants in each session; area of each violin represents the total number of sessions in each condition in which at least one error on both experimentally-restricted and unrestricted consonants was made. Horizontal lines show quantiles. Significance marks are derived via the significance tests described in Section 2.3.2.2.4.

$^{***}p < 0.001$

Figure 7

Consonants' Probability of Preserving Position in Errors Made in Experiment 2



Note. Position-preservation probability was calculated for each session; area of each violin represents the total number of sessions in which at least one error involving a consonant of the respective type-context combination was made. Horizontal lines show quantiles. Significance marks are derived via the significance tests described in Section 2.3.2.2.4.

$^{*}p < 0.05$, $^{***}p < 0.001$

Table 6

Experiment 2 Speech Error Mixed-Effects Model Results

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	0.80	0.13		
session	0.01	0.16	< 0.005	0.944
restricted-constrained vs. restricted-free	0.92	0.33	6.56	< 0.05 *
unrestricted vs. restricted	-1.20	0.21	22.8	< 0.001 ***
similarity to training	-0.97	0.26	12.39	< 0.001 ***
restricted-constrained vs. restricted-free × *session*	0.21	0.47	0.19	0.662
restricted-constrained vs. restricted-free × *similarity to training*	1.93	0.66	8.21	< 0.01 **
unrestricted vs. restricted × *session*	-0.60	0.32	3.43	0.064 ~
unrestricted vs. restricted × *similarity to training*	-0.66	0.42	2.24	0.135
similarity to training × *session*	0.11	0.32	0.11	0.739
restricted-constrained vs. restricted-free × *similarity to training* × *session*	-1.85	0.93	3.56	0.059 ~
unrestricted vs. restricted × *similarity to training* × *session*	-0.14	0.64	0.05	0.832

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4.

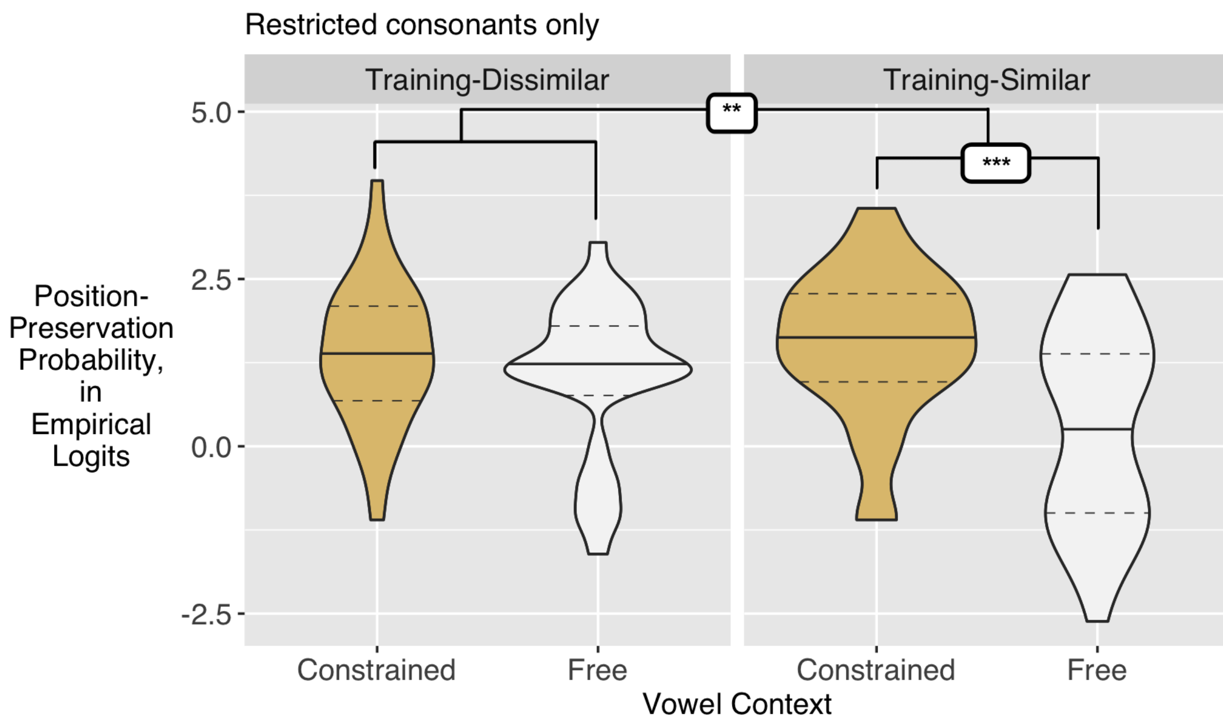
The odds of unrestricted consonants preserving position were approximately 0.30 (β = -1.20, SE = 0.21) times as high as those of restricted consonants preserving position. Restricted consonants' odds of preserving position were 2.51 times higher when they appeared in their constrained context compared to their free context (β = 0.92, SE = 0.33), indicating that participants successfully learned the second-order phonotactic restriction in question.

To test whether there was a significant difference in position-preservation rates between restricted consonants in free contexts and unrestricted consonants, we again re-analyzed our data using a version of the model that set restricted consonants in free contexts as the baseline in a simple coding scheme, as in Experiment 1 (see Appendix B for details of model and results). This analysis revealed a significant difference between unrestricted and restricted consonants in free vowel contexts, with the odds of unrestricted consonants preserving position being 0.52 times those of restricted consonants in free vowel contexts preserving position (β = -0.66, SE = 0.25; $\chi^2(1) = 6.33$, $p < 0.05$). As in Experiment 1, exposing participants to stimuli in which a target consonant was constrained to a certain position in the context of a specific vowel still influenced their errors on that consonant, even in contexts where it was not constrained.

In the main model, we found a significant interaction between the predictors *restricted-constrained vs. restricted-free* and *similarity to training*, such that odds of position-preservation of errors made by participants in the *training-similar* condition on experimentally-restricted consonants in constrained vowel contexts were 6.90 times higher (β = 1.93, SE = 0.66; $\chi^2(1) = 8.21$, $p < 0.01$) than would be expected, in the absence of this interaction. Visual inspection of the data (see Figure 8) revealed that position-preservation probability for experimentally-constrained consonants was comparable across experimental conditions, but errors made by participants in the *training-similar* condition on experimentally-restricted consonants in free contexts were much less likely to preserve syllable position than errors of the same type made by participants in the *training-dissimiliar* condition. To determine whether this was the case, we split the data by condition, and used this data to train a version of the model that included all fixed and random effects except for *condition* and its interactions. The *restricted-constrained vs. restricted free* predictor was significant when this model was trained on data from the *training-similar* condition (β = 1.90, SE = 0.45; $\chi^2(1)$ = 6.68 p < 0.001), but not on data from the *training-dissimilar* condition (β = 0.03, SE = 0.48; $\chi^2(1)$ = 0.004, p = 0.947). Overall, this meant a greater difference between contexts of experimentally-restricted consonants for participants in the *training-similar* condition, compared to those in the *training-dissimilar* condition.

Figure 8

Position-Preservation Probability of Experimentally-Restricted Consonants in Experiment 2



Note. Figure contains data from experimentally-restricted consonants only. Position-preservation probability was calculated within each session. The area of each violin represents the total number of sessions in which at least one error on an experimentally-restricted consonant was made, in a certain vowel context (constrained or free). Horizontal lines show quantiles. ***$p < 0.001$, **$p < 0.01$

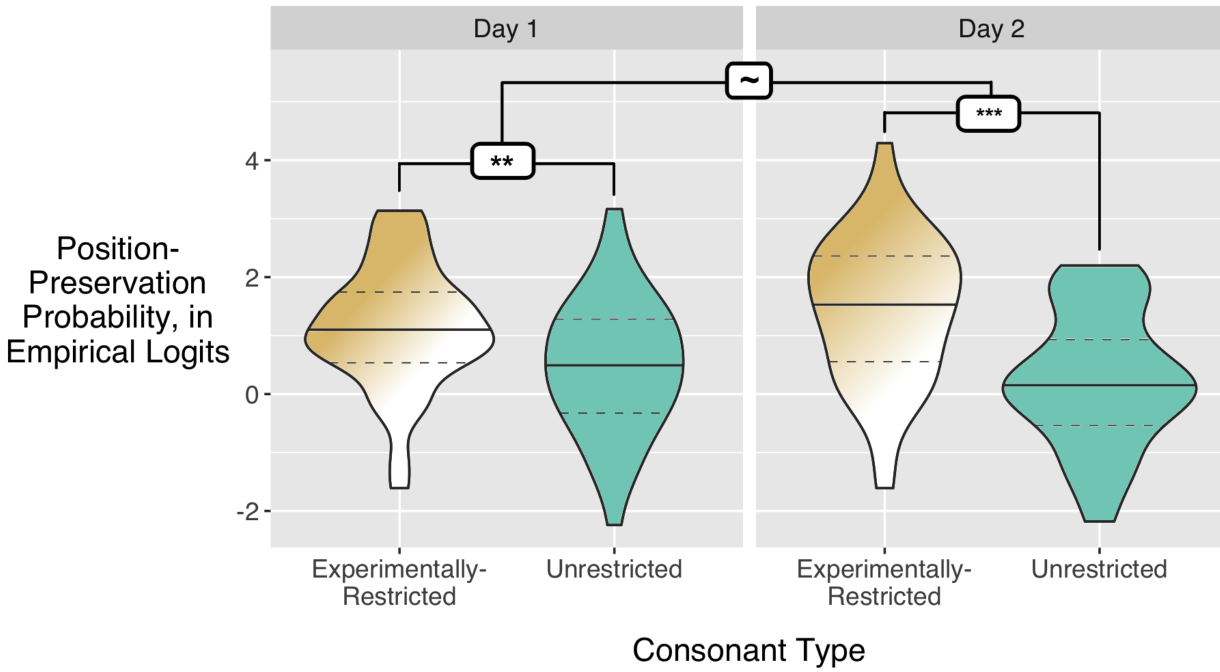
2.4.2.5 Consolidation Effects. We additionally found a marginally significant[9] interaction between the predictors *unrestricted vs. restricted* and *session*, such that the odds of unrestricted consonants preserving syllable position in errors were 0.55 times as high (β = -0.60,

[9] For the purposes of this discussion, we will treat marginally significant effects as reliable.

SE = 0.32) on Day 2 as would be expected in the absence of this interaction. Visual inspection of the data (see Figure 9) suggested that this interaction was driven by a greater difference in position-preservation probability between consonant types on Day 2, compared to Day 1. To verify this, we trained a version of the model that did not include *session* or its interactions on data from each day. The *unrestricted vs. restricted* factor was significant on both days ($\chi^2(1)_{\text{Day 1}} = 9.73$, $p_{\text{Day 1}} < 0.01$; $\chi^2(1)_{\text{Day 2}} = 19.54$, $p_{\text{Day 2}} < 0.001$), but was larger on Day 2 ($\beta = -1.59$, SE = 0.34) than on Day 1 ($\beta = -0.95$, SE = 0.30), meaning that the difference in position-preservation probability between restricted and unrestricted categories was indeed greater on Day 2 than on Day 1.

Figure 9

Position-Preservation Probability by Session and Consonant Type in Experiment 2



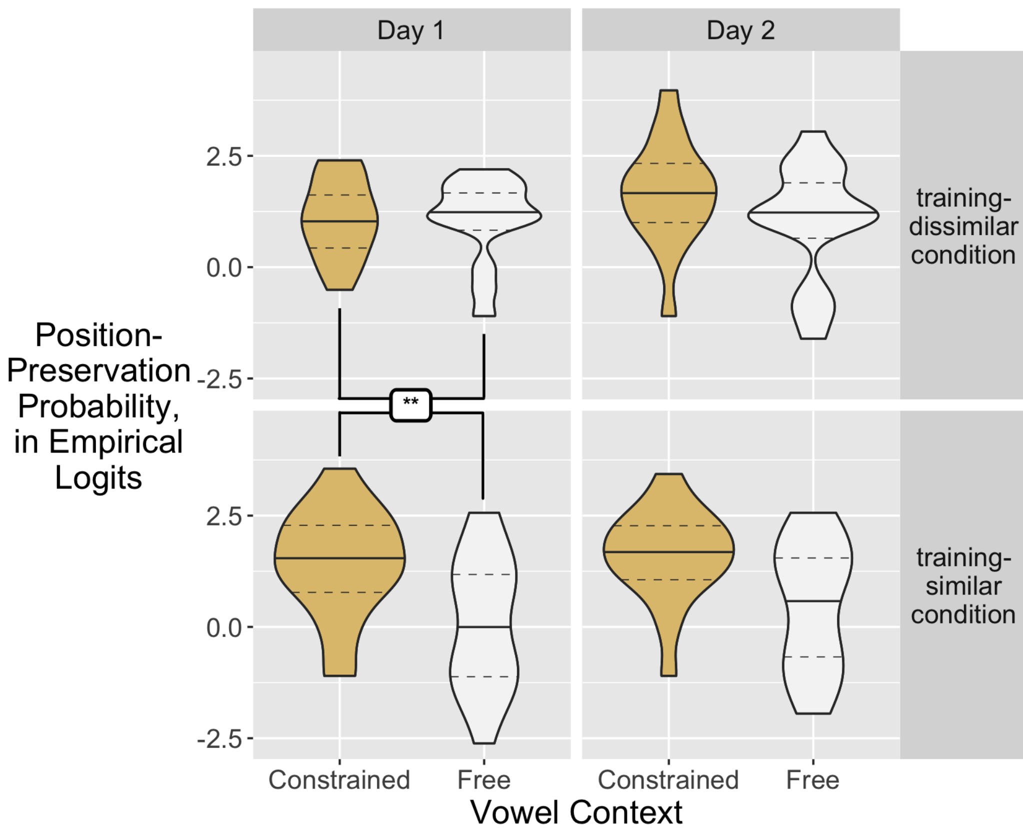
Note. Position-preservation probability was calculated within each session. The area of each violin represents the total number of sessions on a certain day in which at least one error involving a certain consonant type (experimentally-restricted or unrestricted) was made. ***$p < 0.001$, **$p < 0.01$, ~$p < 0.065$

We used these same models to probe the marginally significant third-order interaction between *session*, *restricted-constrained vs. restricted-free*, and *condition* from our main model. The interaction between *restricted-constrained vs. restricted-free* and *condition* was significant on Day 1 ($\beta = 2.76$, SE = 0.93; $\chi^2(1) = 8.53$, $p < 0.01$), but not on Day 2 ($\beta = 1.21$, SE = 0.95;

$\chi^2(1) = 1.75$, $p = 0.186$). This suggests that by Day 2, participants in the *training-dissimilar* condition had overcome some of their interference, in order to attain similar performance as participants in the *training-similar* condition (see Figure 10).

Figure 10

Position-Preservation Probability by Session, Condition, and Vowel Context in Experiment 2



Note. Position-preservation probability was calculated within each session. The area of each violin represents the total number of sessions on a certain day in which at least one error involving a certain consonant type (experimentally-restricted or unrestricted) was made.

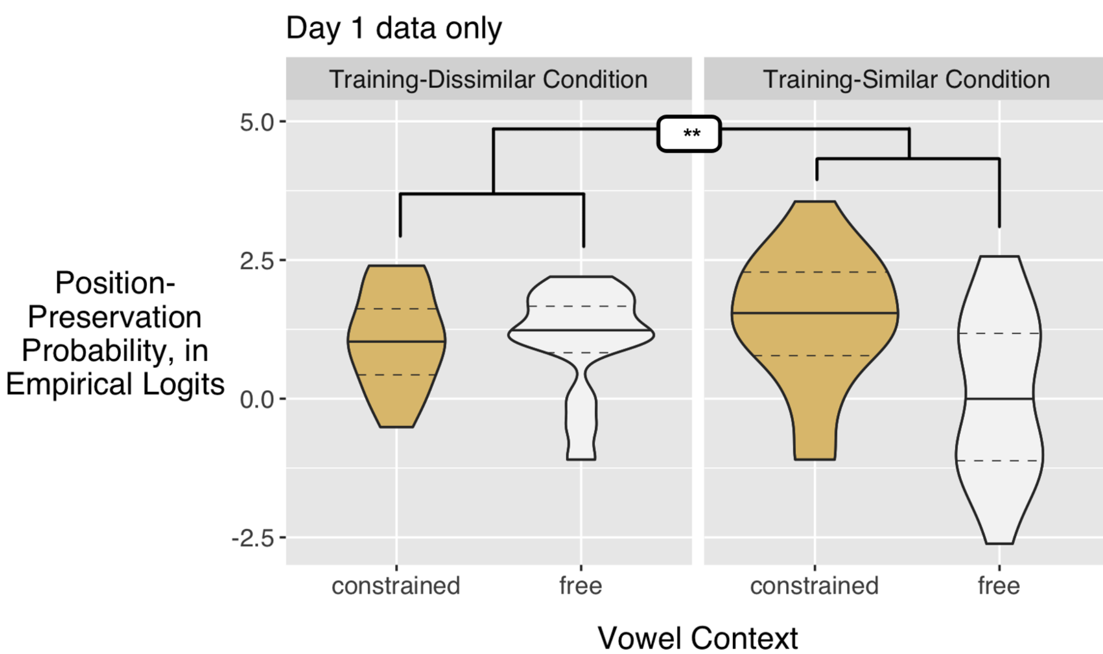
**$p < 0.01$

To attempt to establish whether these effects were tied to consolidation, we constructed another logistic regression predicting position-preservation of errors in which we replaced the *session* predictor from the original model with the interaction between *session* and the z-score corresponding to the proportion of time spent in slow-wave sleep between sessions (*SWS proportion*). Since only 23 participants whose data was included in this experiment also recorded biometric sleep data and provided full demographic information, this comparison was performed only on data from those 23 participants. We then compared the fit of this model to that of the original model (described in Section 2.4.2.2) on the same data, using the likelihood ratio test described in Section 2.3.2.2.4. The addition of slow-wave sleep as a predictor did not significantly improve fit, according to a likelihood ratio test between the slow-wave sleep model and the full model ($\chi^2(1) = 1.96$, $p = 0.161$), so we cannot conclude that this difference between days is due to slow-wave sleep. That being said, the interactions with *session* were also not significant in the baseline (original) model for this 23-participant sample ($\beta = -0.88$, SE = 1.08; $\chi^2(1) = 0.65$, p = 0.421), so it is possible that this follow-up analysis was hampered by a lack of statistical power.

In our follow-up analysis of Day 1 data (see Appendix D for details) we found a significant effect of the *unrestricted vs. restricted* predictor ($\beta = -0.95$, SE = 0.28; $\chi^2(1) = 9.73$, p < 0.01) such that the odds of unrestricted consonants preserving position were about 0.39 times those of restricted consonants preserving position. In contrast to Experiment 1, we did not find a significant main effect of *restricted-constrained vs. restricted-free* on Day 1. Instead, we found a significant interaction between *restricted-constrained vs. restricted-free* and *similarity to* *training*, such that only participants in the *similar-to-training* condition preserved position less often on experimental consonants when they appeared in their unconstrained contexts, as in the main dataset (see Figure 11, compare to Figure 8).

Figure 11

Position-Preservation Probability of Experimentally-Restricted Consonants on Day 1 of Experiment 2



Note. Figure contains data from experimentally-restricted consonants on Day 1 only. Position-preservation probability was calculated within each session. The area of each violin represents the total number of sessions in which at least one error on an experimentally-constrained consonant was made, in a certain vowel context (constrained or free). Horizontal lines show quantiles.

$^{**}p < 0.01$

2.4.2.6 Number of Errors. On average, participants produced 88.97 errors per session (minimum = 12; maximum = 342). In order to test whether the number of errors in each session may have mediated the effects we found in our analysis, we first tested whether the number of errors per session correlated with either *similarity to training* or *session* by fitting a mixed-effects negative binomial linear model that predicted number of errors from *similarity to training*, *session*, and their interaction. Using the model simplification procedures described in Bates et al. (2015), we were able to retain random intercepts by participant in the model. We used the same significance testing procedure described in Section 2.3.2.2.4 to test whether each predictor was significant. Results are shown in Table 7.

Table 7

Experiment 2 Number of Errors Model Results

Predictor	Estimate	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	4.19	0.13		
session	-0.03	0.06	0.29	0.593
similarity to training	0.13	0.25	0.67	0.414
similarity to training × *session*	-0.03	0.12	0.05	0.816

Note. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4.

Since we did not find any significant correlations between number of errors and any of the session-level variables, we conclude that the results of our main model are not likely the result of differences in the number of errors across days or conditions.

2.4.3 *Discussion*

In this experiment, in order to assess whether consolidation can integrate phonotactic knowledge learned in similar contexts, we tested whether the benefits of consolidation for phonotactic learning are modulated by newly-learned phonotactic restrictions' similarity to phonotactic restrictions previously acquired in a similar learning environment. As in Experiment 1, we did not find evidence of this differential benefit. We again found evidence of phonotactic learning in this experiment, with restricted consonants preserving position more often than unrestricted consonants in errors, and restricted consonants preserving position more often when they appeared with their constraining vowels than in their unconstrained contexts. We also replicated the result from Experiment 1 that participants who learned a vowel-coda restriction preserved position more often on errors than those who learned a vowel-onset restriction. In contrast to Experiment 1, we found that being constrained to a specific position boosted position-preservation rate more for consonants produced by participants in the *training-similar* condition, compared to those in the *training-dissimilar* condition, suggesting transfer from pre-training. We additionally found marginally significant evidence that restricted consonants preserved their position more often after a period of sleep, which signals that consolidation may serve to overcome interference stemming from similar learning environments.

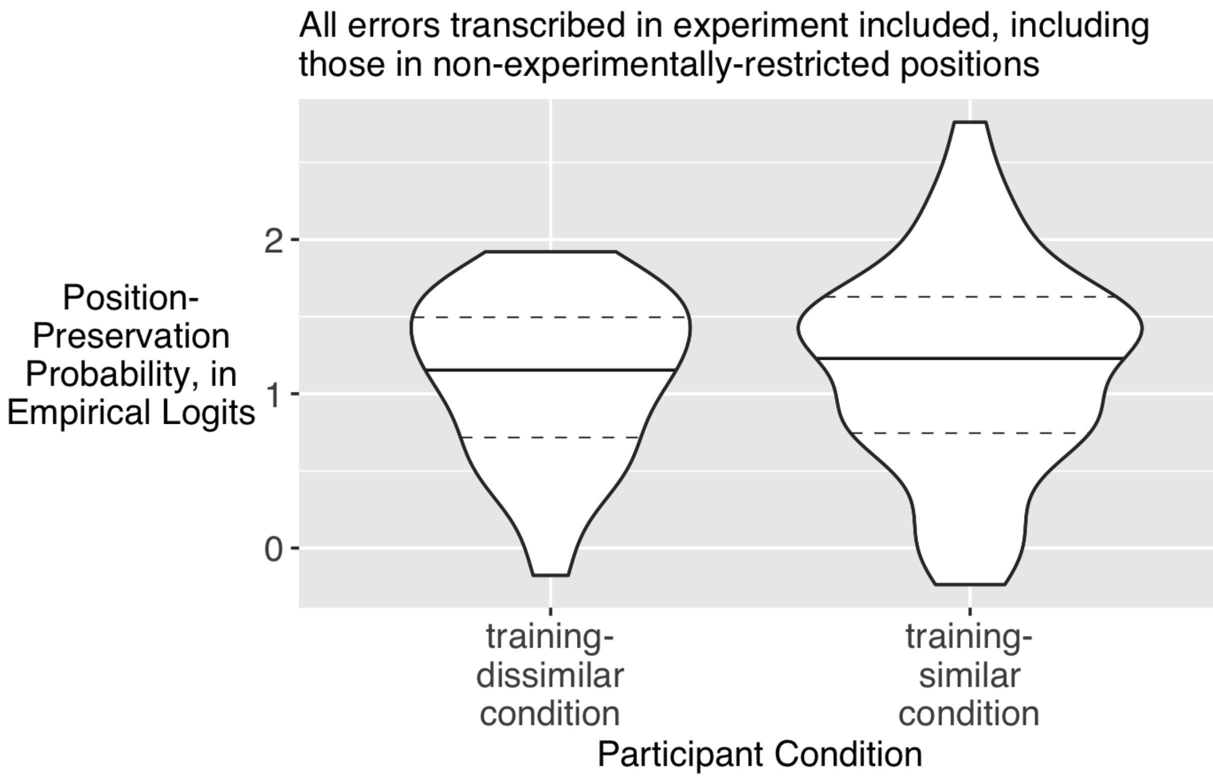
In this experiment, as in Experiment 1, we found that restricted consonants preserved syllable position more often in errors than unrestricted consonants, and that restricted consonants were significantly more likely to preserve position when they appeared in their constrained context. Taken together, these results indicate that participants did use the statistical regularities present in the stimuli to guide their productions.

However, we discovered that the difference in position preservation rates among restricted consonants between constrained and free vowel contexts was only significant for participants in the training-similar condition—hence the interaction between condition and constraining vowel context. This indicates that participants in the training-dissimilar condition did experience interference from the tongue-twisters they recited in the pre-training phase, which contained a vowel-onset phonotactic restriction. The primary effect of this interference in the training-dissimilar condition was to block association of experimentally-restricted consonants with coda. The marginally significant interaction of context and condition with session suggests that training-dissimilar participants could overcome this interference after a period of consolidation. In a follow-up test on data from the present experiment, we did not detect an interaction between this effect and slow-wave sleep, but this may have been due to a considerable reduction in power due to limitations associated with the Fitbits.

In this experiment, errors made by participants in the training-dissimilar condition preserved position more often overall than those in the training-similar condition. Underlyingly, however, this effect resembled the main effect of condition found in Experiment 1, such that participants who were trained on a vowel-coda restriction appeared to make errors that preserved position more often than those trained on a vowel-onset restriction. Again, when we inspect all errors (and not just those made on consonants that appeared in the position targeted by the restriction), we observe that the between-condition difference disappears (see Figure 12). We also again observe that codas tend to preserve position more often than onsets, in both conditions (see Figure 13)—possibly due to English speakers' tendency to keep rhymes intact.

Figure 12

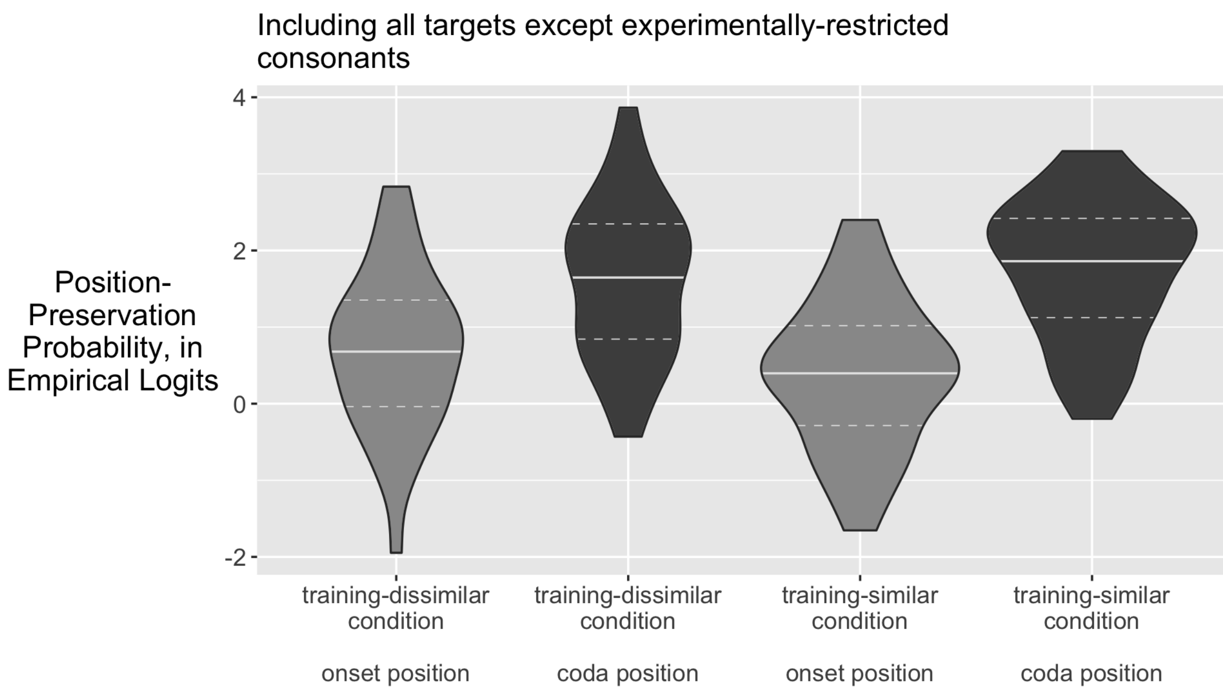
Position-Preservation Rates in Experiment 2, Including All Transcribed Consonants



Note. Area of each violin represents the total number of sessions in each condition. Horizontal lines show quantiles.

Figure 13

Coda Position-Preservation Bias across Non-Experimentally-Restricted Consonants in Experiment 2



Note. Area of each violin represents the total number of sessions in each condition in which at least one error on a consonant in the respective position was made. Horizontal lines show quantiles.

We found a marginally significant interaction between restrictedness and session, such that the difference in position-preservation rate between restricted and unrestricted consonants was greater after a period of sleep-based consolidation. This supports the view that consolidation makes a new, contextually-bound “mini-grammar” (that is separate from the main grammar, yet copied from it) available for syllabification (Dell et al., 2021). In this experiment, consolidation helped participants integrate information about the new consonants seen on Day 1 with their memory representations of consonants from pre-training into a combined mini-grammar. In Experiment 1, on the other hand, participants did not have access to a mini-grammar acquired in a similar context, since the previously-learned phonotactic generalizations of English were acquired naturalistically, during normal childhood language acquisition; this may be why we did not find an interaction between restrictedness and session in that experiment, despite having more statistical power. Taken together, the results of these experiments suggest that the learning environment is important for determining whether new information will be integrated into an existing mini-grammar. In a follow-up test on data from the present experiment, we did not detect an interaction with slow-wave sleep, but this may have been due to a considerable reduction in power due to limitations associated with the Fitbits.

2.5 General Discussion

In this study, we tested whether consolidation may help to integrate old phonotactic knowledge with new, selectively benefiting new phonotactic restrictions that are similar to previously-learned ones. Overall, we found limited support for our hypothesis. Although we found evidence of phonotactic learning in both experiments, we were only able to detect a significant effect of session in Experiment 2, where the environmental contexts in which prior and novel phonotactic restrictions were learned resembled each other. We found additional traces of interference from pre-training in the interaction between condition and constraining vowel context from Experiment 2: despite equivalent power in both conditions, only participants in the *training-similar* condition preserved position significantly more often in errors involving restricted consonants when those consonants were in their constrained vowel context. Moreover, we found that codas in both experiments were overall more likely to preserve position, which we believe may be attributed to the characteristics of English, our participants' native language. Our results are summarized in Table 8.

Table 8

Summary of Speech Experiment Results

Result	Proposed explanation(s) for result
In contrast to previous studies, participants were able to restrict positions of constrained consonants on Day 1 of Experiment 1.	Lack of interference between consonant-position mappings.
In Experiment 2, position-preservation rate difference between restricted and unrestricted consonants was larger after an opportunity for sleep-based consolidation.	Consolidation enabled participants to draw upon relevant prior knowledge from pre-training in a similar learning environment to overcome interference that may have existed on Day 1 of Experiment 2.
In Experiment 2, only participants in the *training-similar* condition consistently preserved position more often on restricted consonants when they appeared in their constrained context, compared to when they appeared in their unconstrained (free) context.	Transfer due to similar learning environment. Interference in the *training-dissimilar* condition due to incongruity between pre-training restrictions and Experiment 2 restrictions.
Experimentally-restricted consonants preserved position more often than unrestricted consonants, even in in vowel contexts that did not constrain their positions.	The frequency of mappings between consonants and syllable positions are represented both locally (with adjacent context) and globally (independent of adjacent sounds) to guide speech production.
Participants preserved position more often on coda consonant errors than onset consonant errors.	Planning differences between the first and third response units. Influence from participants' native language (English), where codas are more tightly bound to vowels than onsets are.

2.5.1 *Consolidation's Roles in Phonotactic Learning*

2.5.1.1 Resolving Competition between Conflicting Positional Mappings. In contrast to many other studies of second-order phonotactic learning, we did not find evidence of a consolidation effect in Experiment 1, despite our large sample size; participants' errors conformed to the phonotactic restriction more often than chance, even on Day 1 (see Appendix D), and our analysis did not indicate a strengthening of this trend on Day 2. We believe this is due to the unique characteristics of our stimuli, and that this difference reveals one of the critical functions of consolidation in previous second-order phonotactic learning studies.

Participants in previous second-order phonotactic learning experiments (Gaskell et al., 2014; Smalle et al., 2017, 2021; Warker, 2013; Warker et al., 2010; Warker & Dell, 2006) learned a set of conflicting restrictions, where each vowel restricted one consonant to onset, and the other to coda. The mapping between restricted consonant and position was totally reversed on trials containing a different vowel. In other words, each vowel constrained the position of both consonants, but this position-consonant mapping was reversed between vowels. Therefore, all associations between a restricted consonant and a given syllable position had to be strengthened in one vowel context, and suppressed in another.[10] Consolidation may serve to overcome this competition among conflicting mappings.

In our experiments, by contrast, the difference between vowel contexts is not so stark: both consonants can appear in the targeted position in any trial, but each consonant is additionally permitted to appear in the untargeted position in stimuli containing the vowel that does not constrain it. Thus, there is less conflict among vowel contexts—the mapping between unconstrained restricted consonant and position targeted by the restriction only needs to be

[10] In Warker and Dell (2006), this property is called "self-interference."

inhibited on 25% of trials for successful (errorless) task performance. It is possible, then, that participants are not engaging inhibitory attentional processes to perform this tongue-twister task (or at the very least, inhibiting less than they would if they were learning two perfectly conflicting restrictions), and therefore can learn the restriction on the first day, without needing consolidation to resolve the competition between mappings. Indeed, we find that in both experiments, participants preserved position significantly more often on errors involving restricted consonants than those involving unrestricted consonants on Day 1, and we did not find evidence that the difference among restricted consonants in constrained contexts and those in unconstrained contexts was greater after a period of consolidation.

This account squares with previous research in which phonotactic restrictions affect speech production on Day 1 in populations with depleted executive function—either through experimental manipulation (Smalle et al., 2021) or age (Smalle et al., 2017). The theory is as follows: in the initial session, attentional control backgrounds seemingly irrelevant information (in this case, the vowel-dependent mappings between consonants and syllable positions) to increase accuracy on the tongue-twister task. During sleep-based consolidation, previously backgrounded information can be foregrounded to improve performance, if that information is relevant to performance on a future task (*c.f.* Stickgold & Walker, 2013), such that the weights of hidden units connecting vowels, consonants, and their restricted positions can be strengthened. In our experiment, hidden units need not be "backgrounded" during the initial session, because these mappings are not in direct conflict with each other, and so the weights can be adjusted online, without the need for an offline consolidation period.

Taken together with prior research, our results suggest that one of consolidation's roles is to resolve interference among mappings between consonants and restricted syllable positions that are in direct conflict.

2.5.1.2 Using Prior Knowledge Acquired in Similar Contexts. Unlike in Experiment 1, we found that the difference in position-preservation rates between unrestricted and restricted consonants was greater after a period of sleep in Experiment 2, suggesting a role for consolidation in phonotactic learning. In Experiment 2, participants had to learn to restrict a new set of consonants (/v/ and /z/) to certain positions in order to correctly produce tongue-twisters. The difficulty associated with learning this new restricted set of consonants was not so strong as to prevent participants from making significantly more position-preserving errors on restricted consonants compared to unrestricted consonants on Day 1 (see Table D1), but the increased difference on Day 2 suggested that consolidation did confer a benefit in distinguishing these two categories of consonants apart for speech production.

What exactly is the nature of this benefit? One possible explanation is that consolidation allowed participants to draw upon their prior experience during the pre-training phase in order to correctly produce tongue-twisters in the new experiment. Under Stickgold & Walker (2013)'s "memory triage" account of consolidation, this is an example of item integration, in which participants integrate newly learned memory representations into pre-existing schemas during consolidation, thus "increasing the schema's potential utility and applicability" (p. 5). For our participants, these new memory representations are the consonants that were encountered during Day 1 of Experiment 2, and the pre-existing schema was the distinction between restricted and unrestricted consonants learned during pre-training (Experiment 1).

The precise neurocognitive mechanism by which this integration occurred is beyond the scope of the current study, but would have important implications for the study of consolidation's role in phonotactic generalization across segments. For example, neuronal replay during consolidation (Lewis & Durrant, 2011) may help participants identify overlapping unrestricted consonants between the two phases of the experiment, thereby permitting the insight that these consonants have the same distribution, regardless of the new consonants in the tongue-twister. Consolidation may also help participants link shared phonological properties (in this case, manner and place of articulation) between restricted consonants in the two phases of the experiment. These possibilities are not exclusive; future research should test them separately.

The fact that we did not find a consolidation effect in Experiment 1 (despite having more statistical power than in Experiment 2) suggests that similarity in the learning/task context facilitates consolidation. This is compatible with Dell et al. (2019)'s proposal that consolidation makes a contextually-bound "mini-grammar" available for speech production. In Experiment 1, participants created an initial experimental mini-grammar, which was siloed off from their experience of English by virtue of its unusual learning environment. This mini-grammar began as a copy of English grammar, but was quickly augmented during the course of the tongue-twister task. In Experiment 2, the experimental mini-grammar was updated to contain the newly presented consonants, and the consolidation that occurred between Days 1 and 2 made this update available for speech production.

It remains to be seen whether this "update" involved the creation of new mini-grammar, specific to Experiment 2 (starting as a copy of Experiment 1), or whether the Experiment 2 consonants were actually integrated into the same experimental mini-grammar as those from Experiment 1—in other words, were the new consonants sufficient contextual cues for the creation of a new mini-grammar? An experiment that tests whether the representations of the restricted consonants from Experiment 1 were modified during Experiment 2 could provide the answer, potentially having implications for the representation of associations between environmental contexts and novel phonotactic restrictions.

2.5.2 *Transfer and Generalization*

In this study, we found that participants transferred experience with a consonant in one context to other contexts and consonants. Restrictions applicable to a certain consonant in the context of a certain vowel were generalized, influencing the distribution of errors on a different consonant (in a different phase of the experiment), as well as to the same consonant in a different vowel context. To our knowledge, these results are the first evidence of transfer in second-order phonotactic learning.

2.5.2.1 Between Consonants, Across Phases. Recall that we found an interaction between participant condition and vowel context among restricted consonants in Experiment 2 (see Figure 8 and discussion of results immediately preceding). Participants exposed to tongue-twisters that followed restrictions similar to those from pre-training made a greater distinction between the constrained and unconstrained contexts of restricted vowels, in terms of position-preservation rates, compared to participants in the *training-dissimilar* condition. This indicates that participants in the *training-similar* condition used their prior knowledge from the pre-training phase to quickly adapt to stimuli containing novel consonants restricted to the same syllable position, as early as Day 1 of Experiment 2 (see Table D2). For participants in the *training-dissimilar* condition, on the other hand, pre-training may have interfered with the learning of new phonotactic generalizations; a follow-up analysis on data from only this group revealed no significant difference in position-preservation rate of errors on restricted consonants between constrained and unconstrained vowel contexts. We also did not find evidence that this difference among conditions lessened after a night of sleep-based consolidation, suggesting that consolidation could not overcome the *training-dissimilar* group's difficulties (though we leave open the possibility that we lacked the statistical power necessary to detect such a change).

The results from the *training-dissimilar* group are consistent with Anderson et al. (2019)'s findings regarding reversal shift in phonotactic learning, which in turn support an incremental account of phonotactic learning. In their single-session tongue-twister study, participants' speech errors exhibited reduced/slowed adaptation to a first-order phonotactic restriction that was presented after participants had experienced 32 trials of its "reversed" version, which inverted the positions to which experimentally-restricted vowels were constrained. This "reversal" is analogous to the shift that our *training-dissimilar* participants experienced between pre-training and Experiment 2: consonants previously restricted to coda were now restricted to onset, and vice versa. The present study adds to our understanding of how phonotactic restrictions are represented in the mind by establishing that second-order phonotactic learning is also incremental: old associations between features of restricted consonants and their constrained positions must be "unlearned" before new ones can affect production. Additionally, our results suggest that this incrementality can unfold over an extended period of several days: the opportunity to consolidate between and within phases does not create a representation that allows the rapid reversal typical of category learning.

At least two factors may have promoted transfer between phases in the *training-similar* condition: (1) phonological closeness between the sets of experimentally-restricted consonants and (2) identical vowel contexts restricting each consonant. Future research will need to establish the contribution of each of these factors, in order to determine the basis of phonotactic transfer.

2.5.2.2 Within Consonants, Across Vowel Contexts. In both experiments, we found that errors on experimentally-restricted consonants in vowel contexts that did not constrain their positions still preserved position more often than errors on unrestricted consonants. Although vowel context did affect the rate of position preservation among errors on experimentally-restricted consonants, participants did not treat experimentally-restricted consonants in unconstrained vowel contexts as equivalent to unrestricted consonants (as they should, if their productions veridically followed the statistical properties of the input).

This result provides support for simultaneous learning of global, first-order probabilistic phonotactic generalizations (e.g., "/f/ appears in coda position 75% of the time, throughout the experiment") and local, second-order phonotactic generalizations (e.g., "/f/ is more likely to appear in coda position, if /ɪ/ is the vowel in this syllable"), such that these generalizations overlap to affect production. With respect to Warker and Dell (2006)'s syllabification model, this modification of the training data allows for ambiguity of analysis, such that multiple statistical generalizations are consistent with the new data. Since our restrictions are not self-interfering (see discussion in Section 2.5.1.1), the system learns to downweight vowel identity as a predictor, compared to a model trained on self-interfering data (as in previous experiments). However, since the identity of a specific vowel is highly correlated with activation of a certain phoneme-position combination, the system can never totally ignore vowel identity. The fact that this result obtained even after an opportunity for sleep-based consolidation suggests that, if an architecture like Warker and Dell (2006)'s is operant in this learning process, there are either not enough hidden units to represent, for each consonant, each vowel context completely separately from the other, or more training is required to overcome the effects of the added ambiguous training data.

2.5.3 *Native Language Influence on Speech Production*

Across both experiments, errors were more likely to preserve positions of consonants that were originally in coda, compared to consonants that originally appeared in onset position. This was true for experimentally-restricted consonants, as well as unrestricted and language-restricted consonants, whose syllable positions were not manipulated across conditions in either experiment. It is unclear exactly why this effect arises, but it may be due to a general property of English: since codas are more predictable from vowels than onsets (Kessler & Treiman, 1997), there may be a tendency to keep the rhyme unit together in speech errors. If this is the source of the bias, we might expect a bias in the opposite direction (onsets more likely to preserve position than codas in errors) for native speakers of a language in which onsets are more predictable than codas, like Korean (Lee & Goldrick, 2008). Smalle and Szmalec (2021) have documented differences in second-order phonotactic learning in adults based on native language: errors from native speakers of French conform to second-order phonotactic constraints before a consolidation opportunity. It is possible that other differences by native language may be detectable in this task, especially since the vast majority of studies using this paradigm have used native English speakers as participants.

On the other hand, this effect may also arise from mere temporal differences: participants have more time to plan codas before producing them, so they have more time to select the appropriate syllable position before the sound must be produced. Chapter 3 in this dissertation will examine and test this hypothesis.

2.5.4 *Improving Data from the Tongue-Twister Paradigm*

We believe there is considerable potential for methodological improvement in the tongue-twister paradigm. First, we recommend that other researchers using this paradigm report agreement statistics for errors in particular (since only errors are analyzed in these experiments), instead of merely an aggregate over the entire dataset. Although overall agreement between transcribers was quite high (see Sections 2.3.2.3.1 and 2.4.2.3), agreement on errors was only 55.54% (Cohen's $\kappa = 0.49$) across both experiments, indicating only fair-to-moderate agreement; we suspect this may be the case for other studies as well, but since statistics on errors are not usually reported separately, we have limited basis for comparison. This comparison across studies is critical for establishing the limitations of this paradigm in detecting errors, under various conditions (transcribers, recording conditions, etc.).

Secondly, we recommend the use of a standardized, blind transcription procedure and an automated error-detection algorithm (such as the one we describe in Section 2.3.2.2.1), in order to minimize the effects of researcher bias. We additionally recommend making analysis scripts, transcriptions, and audio files (with participants' permission) available for future analysis by interested researchers. Open data of this type will allow re-transcription of data for reliability, encourage development of new error-detection algorithms, and provide a training set for automated speech-to-text software that can greatly improve the practicality of this research.

Last but not least, we advocate the development of automated speech-to-text software that is specialized for this task. Although the web-based data collection software we developed can quickly capture recordings from a large participant sample, transcription remains a manual process, since current speech-to-text software generally aims to minimize disruption due to errors in speech production instead of registering errors for further analysis. Thus, transcription is presently the most time- and labor-intensive step of conducting research using the tongue-twister paradigm. This high barrier to entry prevents all but a few groups from using this task to study phonotactic learning. In this environment, progress is slow, and replication studies are difficult. Automated transcription would address these issues.

2.6 Conclusion

In this study, we investigated prior knowledge's role in second-order phonotactic learning, along with consolidation's potential moderation of that role, using a tongue-twister task across multiple days. We found that the opportunity for consolidation allowed participants to use prior knowledge acquired in a similar context to help them produce speech following novel phonotactic restrictions. Our results indicate that participants needed to unlearn the mapping between features of restricted consonants and the syllable positions to which they were constrained before a mapping to the opposite syllable position could affect their productions; the rapid reversal typical of abstract category learning was absent, even after a period of consolidation. Additionally, we find that probabilistic first-order phonotactic restrictions are learned simultaneously with second-order restrictions; global probabilities of restricted consonants appearing in constrained syllable positions bias participants' responses to favor those syllable positions even in contexts where those restricted consonants are unconstrained. Taken together with prior work, this study's results support a view of second-order phonotactic learning that is incremental and probabilistic across multiple levels of representation. Consolidation, in turn, helps participants transfer helpful prior knowledge acquired in similar contexts of learning and resolve interference arising from sets of phonotactic restrictions that categorically map the phonetically similar consonants to different syllable positions.

Learning Sequential Patterns in the Absence of Structured Prior Knowledge

3.1 Introduction

In Chapter 2, we demonstrated that consolidation's function in phonotactic learning is modulated by the prior phonotactic knowledge that participants have about the task—its content, as well as the context in which the knowledge was acquired. The role of consolidation in the absence of prior knowledge, however, is underdetermined. To examine this, we attempt to conceptually replicate Experiment 1 in a production modality where participants have little, if any, prior knowledge regarding sequential regularities: a button-pressing task.

3.2 Learning Regularities in Keypress Sequence Production

3.2.1 *Analogy to Speech Production*

Speech errors have a tendency to follow phonotactic generalizations in a speaker's language, whether these speech errors occur naturalistically (Boomer & Laver, 1968; Fromkin, 1971; Motley, 1973) or in the lab (Motley & Baars, 1975). Studies of phonotactic learning for speech production have leveraged this property to infer the current state of speaker's phonotactic knowledge, under the assumption that participants' errors should follow a phonotactic restriction more often than expected by chance, if that phonotactic restriction has been learned.

But sequential regularities are not exclusive to the domain of language; phonotactic regularities are only a subset of the sequential patterns that exist throughout the natural world. There is evidence that the same domain-general capacities may be employed when learning sequential regularities in production, regardless of whether that production occurs via speech or a different response modality. In particular, two recent studies (Anderson & Dell, 2018; Rebei et al., 2019) found similarities between sequential regularity learning in speech production and learning in a button-pressing task. Both studies found that a consolidation opportunity in the form of sleep enhances the tendency of errors to follow the sequential regularities present in the stimuli of the experiment, but only for sequential regularities that involve the coordination of multiple response units (phonemes for speech production, keypresses for button-pressing)—i.e., second-order sequential regularities. Moreover, Anderson and Dell (2018) found that first-order constraints, which restrict a single response unit, can affect errors in button-pressing without the need for a consolidation period—just as in speech production (Dell et al., 2000). Additionally, Anderson et al. (2019) examined reversal shift in both speech production and button-pressing, and found that reversed regularities are learned more slowly than initially-learned regularities in both modalities, suggesting that the statistical learning process is incremental, regardless of response modality.

Anderson and Dell (2018) and Rebei et al. (2019) also find critical dissimilarities between button-pressing and speech production, which they identify as arising from differences in structured prior knowledge across modalities. Both studies find that errors preserve position less often overall in the button-pressing task than they do in speech production—particularly for unrestricted consonants. In addition, though improvement was found after a night of sleep, both studies found evidence of second-order sequential regularities affecting production on Day 1. In Anderson and Dell (2018), restricted finger key targets preserved position in 50% of errors, and unrestricted targets preserved position in 40% of errors, leading to a 1.51 odds ratio between restricted and unrestricted. This odds' ratio was 1.27 for Rebei et al. (2019). The authors attribute the differences between response modalities to a weaker yet more flexible schema for production of button-presses, similar to the weaker yet more flexible schema displayed by children (Smalle et al., 2017).

3.2.2 *Current Study: Structured Prior Knowledge (or Lack Thereof)*

In the present experiment we ask whether the effects from Experiment 1 of Chapter 2 are domain-general or specific to language. These effects, along with their predictions for the current experiment, are summarized in Table 9. Taken together, the results of the experiments in Chapter 2 suggest that consolidation serves to resolve interference among conflicting mappings between response units (consonants) and temporal (syllable) positions, and that consolidation allows participants to use prior knowledge acquired in similar contexts to guide task performance. We have reason to believe that some of these results may arise from the demands of a production task, rather than as a consequence of prior linguistic knowledge.

Table 9

Summary of Speech Experiment Results (Repeated from Table 8) and Predictions for Button-Pressing Experiment

Result from speech experiments	Proposed explanation(s) for result	Prediction (if phonotactic learning results are domain-general)
In contrast to previous studies, participants were able to restrict positions of constrained consonants on Day 1.	Lack of interference between consonant-position mappings.	Similar; participants should be able to restrict the position of finger keys targeted by the restriction on Day 1.
Experimentally-restricted consonants preserved position more often than unrestricted consonants, even in in vowel contexts that did not constrain their positions.	Mappings between response units and their positions are represented both locally (with adjacent context) and globally (independent of adjacent response units).	Similar; restricted finger key targets in unrestricted (free) contexts will preserve position more often than unrestricted consonants.
Participants preserved position more often on coda consonant errors than onset consonant errors.	Planning differences between the first and third response units.	Similar; third-key errors will preserve position more often than first-key errors.

In Chapter 2, we hypothesized that the reason participants were able to restrict positions of constrained consonants on Day 1 of Experiment 1 was that our mappings between restricted consonants and their syllable positions were not self-interfering (see Row 1 of Table 9). The restrictions could therefore influence participants' speech errors on Day 1, without needing consolidation to reconcile the interference stemming from inhibition of conflicting mappings. In the button-pressing experiment, our stimuli are similarly non-interfering. If our findings from phonotactic learning are solely driven by properties of general sequential pattern learning, we expect the same result: participants should be able to restrict the position of response units (in this case, finger keys) targeted by the restriction in errors on Day 1. If, however, the lack of interference in Experiment 1 was a consequence of the "strong yet rigid" syllable schemas that adults have acquired over a lifetime of speech production, we expect to find limited learning of the restriction on Day 1, when no pre-existing schema exists to guide keypress sequence production; the restriction may only influence speech errors on Day 2, once consolidation has resolved issues arising from needing to inhibit responses in this new task.

In both experiments of Chapter 2, we found that, although experimentally-restricted consonants in constrained context preserve position most often overall, experimentally-restricted consonants preserved position more often than unrestricted consonants, even in vowel contexts that did not constrain their positions (Row 2 of Table 9). This may mean that participants represent second-order, phonotactic restrictions simultaneously with first-order, global phonotactic generalizations, and that both of these overlap to influence speech production. If this phenomenon generalizes to sequence production in other modalities, such that sequential patterns are simultaneously represented at the local (context-dependent) and global (context-independent) levels, we expect participants' errors in a button-pressing task to also follow this pattern. Specifically, restricted finger key targets should appear in their constrained position more often than unrestricted finger key targets, even in trial contexts where their distribution is unconstrained. If, on the other hand, participants represent phonotactic generalizations this way because it is specifically helpful for the linguistic system, we might expect them to represent the distributions of finger key targets more veridically in this task, and avoid generalizing restricted consonants' distributions from constrained to the unconstrained contexts.

Additionally, we found that coda consonants preserved position significantly more often than onset consonants in both experiments in Chapter 2 (see Row 3 of Table 9). This may be due to the tight relationship between vowels and codas in English (Kessler & Treiman, 1997), coupled with our English-speaking participants' internalization of that tight relationship in their syllable-production schema. Another possibility is that the effect stems from the nature of sequential planning processes more generally, and has nothing to do with language per se. If the effect is also present in keypress sequence production, with errors involving the third target key in a subsequence preserving position more often than errors involving the first target key, it would serve as evidence that the effect is due to task demands, instead of English linguistic knowledge.

3.3 Method

3.3.1 *Participants*

We recruited 51 adult native English speakers with no self-reported history of speech, language, hearing, or sleep disorder or color vision impairment through flyers posted on the Northwestern University campus, as well as via the Northwestern University Psychology Department's online list of paid research opportunities, which was accessible to the general public. Each participant completed a screening questionnaire ahead of participation to establish that they met the aforementioned demographic criteria, as well as to ensure that they had access to the equipment required to complete the study: a computer with a USB port and headphones. The screening survey also contained questions about availability; participants were only eligible if they could be available for 1 hour at the same time of day for two weekdays in a row.

All participants provided informed consent before taking part in the study. Participants were compensated a total of $30 for both days: $13 for each day, plus an additional $4 bonus payment for completing both days to incentivize return.

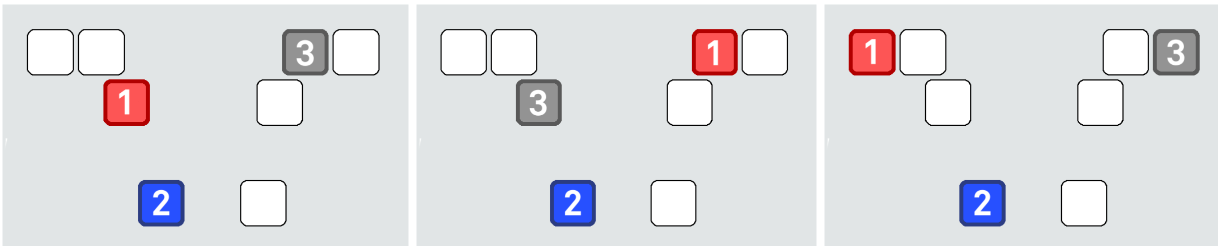
A total of 19 participants were excluded from analysis due to experiment software malfunction ($N = 2$), failing to complete at least one session ($N = 3$), attrition between days ($N = 2$), not making at least one error on both a restricted and an unrestricted key ($N = 3$), taking more than two hours to complete either session ($N = 4$), producing either more than 40 ($N = 1$) or less than 36 ($N = 3$) keypresses on over 50% of trials (indicating that participants may have misunderstood the task), or producing more than 18 errors on over 20% of trials (suggesting either an experiment malfunction or a misunderstanding of the experimental task; $N = 5$), for a final sample size of 32 participants.

3.3.2 *Materials*

Each stimulus in the button-pressing task consisted of a tableau with three panes (see Figure 14 for an example). Each pane represented a subsequence of keys that participants were asked to press in a specific order: the first key was colored red, the second key was colored gray, and the last key was colored blue.

Figure 14

Example of Stimulus Tableau With Three Panes



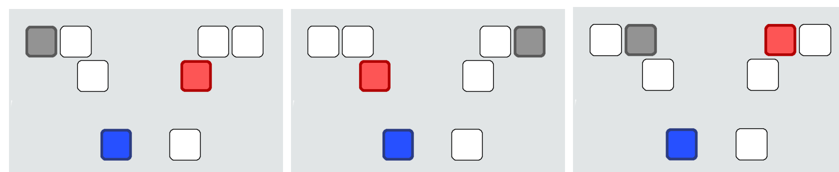
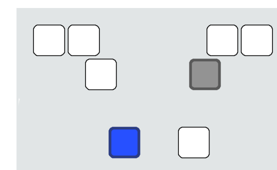
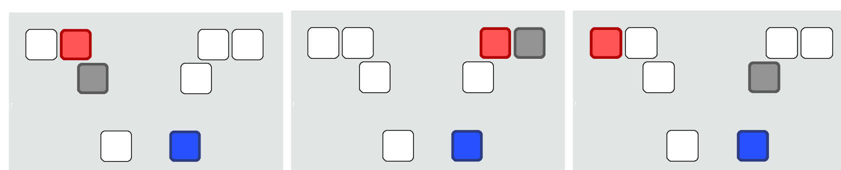
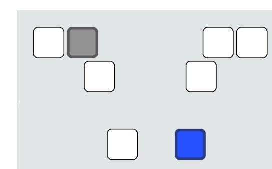
Note. Order numbers are included here for clarity, but in the experiment, participants did not see these numbers; color was the only cue to order. All participants were asked to press the keys in red-blue-gray order.

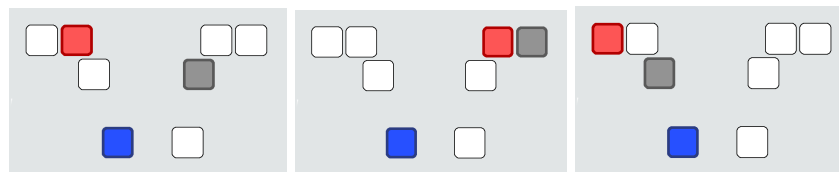
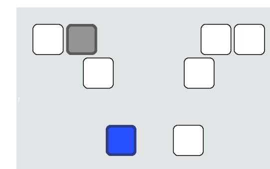
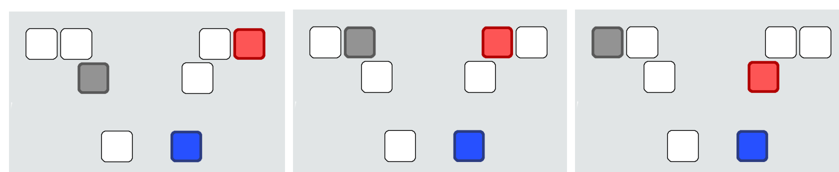
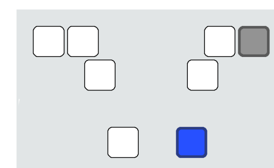
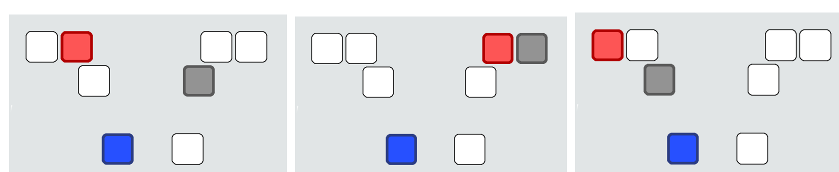
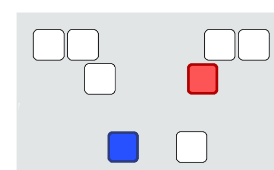
Sequence Structure. The first and third keys always corresponded to non-thumb keys ("finger keys" hereafter), while the second key was always a thumb, reproducing the structure of previous button-pressing experiments that established analogies to phonotactic restrictions (Anderson & Dell, 2018; Rebei et al., 2019). In these experiments (as in ours), thumb keys are analogous to vowels, and finger keys are analogous to consonants. Within each tableau, the same thumb was used in all three panes in order to encourage errors, and each finger key appeared exactly once to allow detection of errors on these keys, which were critical to our dependent measure. Half of the tongue-twisters that each participant saw contained the left thumb, and half contained the right thumb.

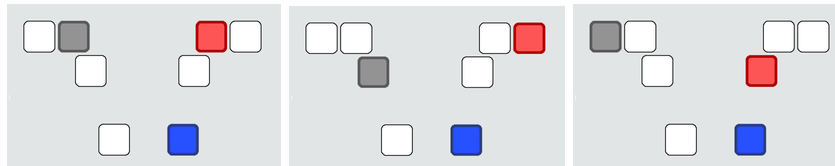
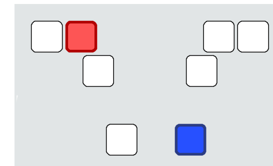
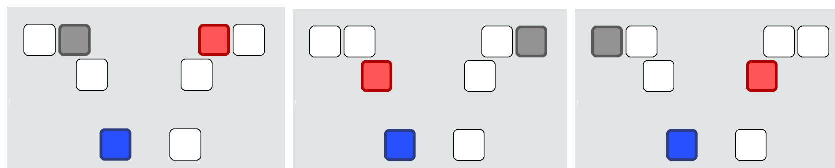
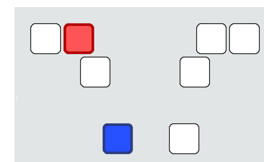
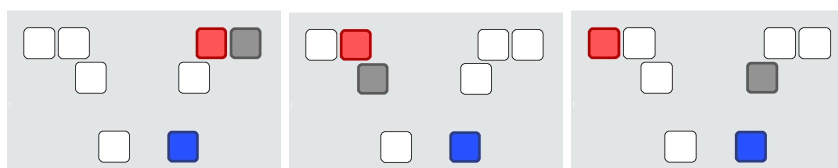
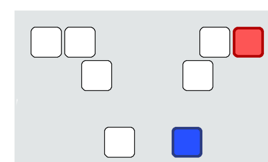
Phonotactic Dependencies. For each participant, two randomly-selected finger keys from different hands were designated as experimentally-restricted, as in Anderson & Dell (2018) and Rebei et al. (2019). These experimentally-restricted consonants were randomly mapped to a thumb key (either left or right) that restricted its position. For participants in the *restricted-to-first* condition, each experimentally-restricted key was constrained to the first position within a pane, in stimuli containing its respective restricting thumb key. Participants in the *restricted-to-last* condition, on the other hand, saw stimuli in which each experimentally-restricted key was constrained to the last position within a pane, in stimuli containing the key's restricting thumb key.

Table 10

Examples of Button-Pressing Stimuli in Each Condition

Experimental condition	Restrictions (restricting thumb → restricted finger)	Example stimulus: Thumb key	Example stimulus: Panes + Finger Key Target Types	Illegal subsequence
restricted-to-first	L thumb → R index; R thumb → L middle	L	 ● restricted-constrained: R index ● restricted-free: L middle ● unrestricted: L ring, L index, R middle, R ring	
		R	 ● restricted-constrained: L middle ● restricted-free: R index ● unrestricted: L ring, L index, R middle, R ring	

Experimental condition	Restrictions (restricting thumb → restricted finger)	Example stimulus: Thumb key	Example stimulus: Panes + Finger Key Target Types	Illegal subsequence
	L thumb → L middle; R thumb → R ring	L	 • restricted-constrained: L middle • restricted-free: R ring • unrestricted: L ring, L index, R index, R middle	
		R	 • restricted-constrained: R ring • restricted-free: L middle • unrestricted: L ring, L index, R index, R middle	
restricted-to-last	L thumb → R index; R thumb → L middle	L	 • restricted-constrained: R index • restricted-free: L middle • unrestricted: L ring, L index, R middle, R ring	

Experimental condition	Restrictions (restricting thumb → restricted finger)	Example stimulus		Illegal subsequence
		Thumb key	Panes + Finger Key Target Types	
		R	 • restricted-constrained: L middle • restricted-free: R index • unrestricted: L ring, L index, R middle, R ring	
	L thumb → L middle; R thumb → R ring	L	 • restricted-constrained: L middle • restricted-free: R ring • unrestricted: L ring, L index, R index R middle	
		R	 • restricted-constrained: R ring • restricted-free: L middle • unrestricted: L ring, L index, R index R middle	

Note. Restriction mappings are non-exhaustive; only two out of 18 possible mappings per condition are displayed here. Bulleted lists identifying restricted consonants in each stimulus are included for illustrative purposes only and were not shown to participants. Likewise, panes demonstrating the phonotactic generalization in question appear first in each condition here for clarity, but in the experiment itself, these panes could appear anywhere in the stimulus tableaux. Illegal subsequences never appeared in stimuli shown to participants in the respective condition.

To construct stimulus lists for each condition, we first generated all 720 possible sequences using these restrictions in each condition. In order to keep this experiment to 97 trials[11] (comparable to the length of our speech experiments), we randomly sampled 194 tongue-twisters from each condition. Since the phonotactic generalization learned by each participant was the same on both days, we randomly sampled 97 tongue-twisters to present on Day 1 for each participant, and presented the remaining 96 tongue-twisters on Day 2.

3.3.3 *Procedure*

Upon submitting the online screening survey, participants were emailed a link to the consent form for the study. After providing their informed consent, participants booked appointments to pick up and return the study equipment in person at the Northwestern University Evanston campus and optionally provided their height, weight, age, and gender for use in the activity tracker's sleep staging algorithm. Pick-up and drop-off appointments were scheduled at least two days apart to allow participants time to complete all experimental sessions between appointments. At the pick-up appointment, the experimenter distributed the study equipment (modified USB keyboard, USB-C adapter, and activity monitor), told participants to begin each experimental session 24 hours apart, and instructed participants to wear the activity monitor while sleeping between sessions. The experimenter then sent each participant a link to the first session of the study no later than 30 minutes after the pick-up appointment. All experimental sessions were completed online.

Due to the COVID-19 pandemic, we used participants' home computers to display stimuli and record responses, using a custom-built web application hosted on a Google Firebase

[11] We initially intended to have 96 trials per session, but due to an error in programming the experimental software and creating the stimulus lists, each session had 97 trials instead.

server.[12] We instructed participants to use either Chrome, Firefox, or Microsoft Edge to access the study. To record responses, we provided participants with a homemade “button box” (see Figure 15): a standard USB QWERTY keyboard that we modified to minimize keypresses on keys that did not appear in the experiment, prevent participants from needing to reposition their hands during the experiment, and defamiliarize the keyboard so that participants would not gravitate toward the “home row” of keys commonly used in touch typing. Modifications were as follows:

1. We affixed green stickers and adhered a circular plastic bump dot over all target keys in the experiment.
2. We removed the top row of keycaps (i.e., those from the function keys) to prevent accidental activation from participants’ wrists during the experiment.

[12] Code for this application is available at https://github.com/nimirea/button-pressing-online-firebase.

Figure 15

Modified USB Keyboard for Recording Keypresses

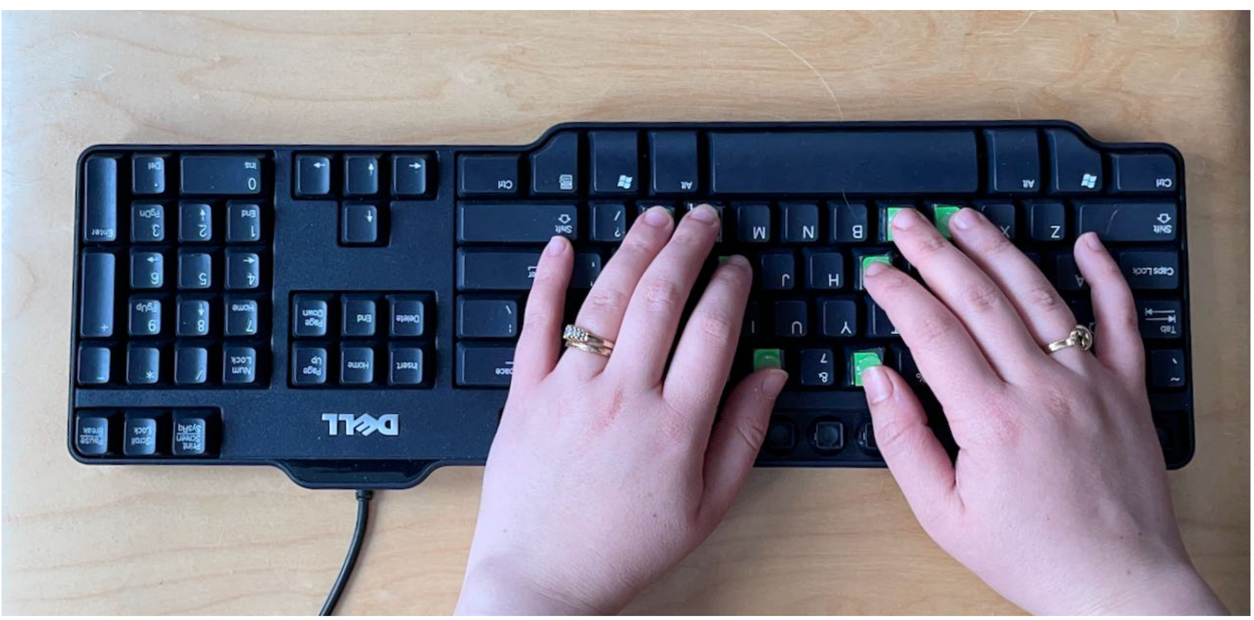
Note. Participants were instructed to plug the keyboard into their computer's USB port and turn it upside-down, so that the space bar was facing away from them, as in this image.

Although these USB keyboards' temporal resolution was slightly lower than that of the Cedrus RB-840 response pads used in previous phonotactic button-pressing studies (10 milliseconds vs. 2-3 milliseconds; *Cedrus Response Pad RB-840*, n.d.), these keyboards had a critical user-friendliness advantage, rendering them more suitable for remote data collection: all participants successfully connected the keyboards to their home computers without needing to install any additional firmware, drivers, or other software. Their low cost (approximately $15 per keyboard, compared to $644.40 for the RB-840) also allowed us to collect data from multiple participants simultaneously, greatly expediting data collection at low risk.

Each experimental session lasted approximately 60 minutes, and sessions were gated such that participants could begin Day 2 no sooner than 24 hours after having begun Day 1. Participants were randomly assigned to a condition (*restricted-to-first* or *restricted-to-last*) with equal probability upon loading Day 1's session. Each session started with an interactive wizard to verify that the participants' keyboard was plugged in and positioned with the space bar facing away from the participant. Participants were instructed to connect their headphones and place their hands on the stickered keys as shown in Figure 16. They were also instructed not to remove their hands from these keys for the entirety of the experiment, and proceeded through the task instructions by pressing both thumb keys simultaneously to navigate. In order to familiarize participants with the instructions and trial structure, each participant completed two guided practice trials at the start of Day 1, and another practice trial at the start of Day 2.

Figure 16

Hand Placement Model for Button-Pressing Task



3.3.3.1 ***Trial Structure.*** At the start of each trial, a fixation cross appeared on the screen, accompanied by a low-pitched (703 Hz) metronome that began to play at a 1 beat per second to signal the beginning of the trial. On the second metronome beat, the fixation cross was replaced with the stimulus tableau. On the fourth beat, participants heard a single medium-pitched (881 Hz) "get ready" beep. On the fifth beat, the metronome increased in pitch again to 1049 Hz (to distinguish it from the "get ready" beep), and the first pane in the tableau was highlighted with a yellow border to indicate that it was time to key it in. Participants were instructed to press one key per beat in each pane from left to right, leaving a one-beat pause between panes. Each fourth beat was higher pitched (1486 Hz) to indicate that participants should pause between panes. Each pane was highlighted by a yellow border during the three beats allotted for it. After 12 beats from the start of the sequence (one complete pass through the 9 keypresses of the tableau, plus three "pause" beats between panes), the metronome sped up to a rate of 2 beats per second, and participants were instructed to repeat the sequence three times at this faster rate. After the 36 fast beats (9 keypresses + 3 pause beats per repetition, times three repetitions), participants received instructions to press both thumbs simultaneously to continue to the next trial. This allowed participants to take self-timed breaks if necessary.

3.3.3.2 ***Surveys.*** At the end of each experimental session, participants were asked what equipment they used for the session (browser, computer operating system, and headphones) to help diagnose data quality issues during data collection. They were also asked how many hours of sleep they received the previous night, and had an opportunity to provide other written feedback about their experience. The Karolinska Sleep Questionnaire (KSQ; Westerlund et al., 2014) was additionally administered after Day 1's session.

3.3.3.3 ***Sleep Data Collection.*** We collected sleep data using Fitbit Inspire 2 wrist-worn activity monitors, which use body movement, heart rate, and heart rate variability to detect sleep/wake intervals and identify different sleep stages (Stage 1, Stage 2, and Deep [Slow-Wave] Sleep). When measured against an EEG machine, Fitbit's algorithms using these variables are comparable to a standard wrist-worn actigraphy device in terms of identifying slow-wave sleep and detecting sleep intervals (Haghayegh et al., 2020). Participants were asked to wear these activity monitors on nights between sessions, on their non-dominant hand, one finger-width up the wrist, as recommended by Fitbit's support website (Fitbit, Inc., n.d.). Before the pick-up appointment, the experimenter created a Fitbit account for each participant using the provided demographic data and paired one of the activity monitors to the account. After drop-off, the experimenter synchronized the activity monitor with the account and copied sleep times, wake times, and slow-wave sleep duration for all recorded sleep sessions to a CSV file.

3.4 Results

3.4.1 *Button-Pressing Error Analysis*

3.4.1.1 Statistical Model. To test the hypothesis that planning effects make the final element in a subsequence more likely to preserve position, even when no structured prior knowledge exists, we built a mixed-effects logistic regression that predicted position-preservation of finger keys pressed in participants' errors using session and participant condition, as well as each pressed finger key's context of appearance and restrictedness. We used the lme4 package (Bates, Mächler, et al., 2015) in the R programming language (R Core Team, 2022) to fit this model.

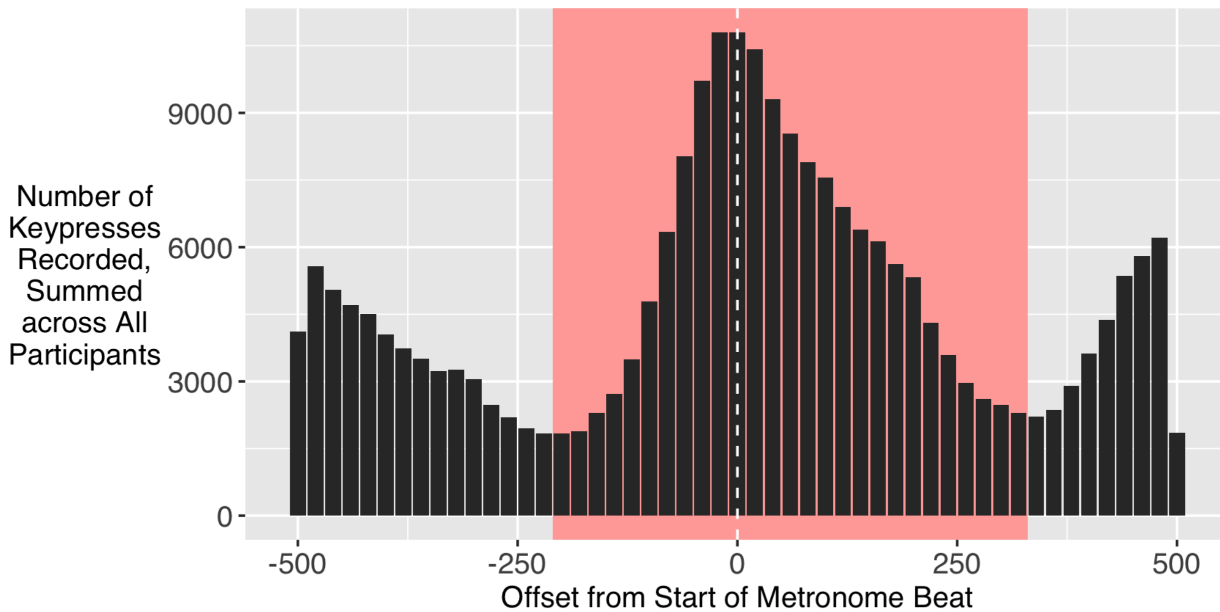
3.4.1.1.1 ***Dependent Measure.*** Our learning measure of interest was the rate at which keys pressed in errors preserved their position within a pane. In our logistic regression, then, the dependent measure was whether the position of each key pressed in error was the same as the position it had in the pane where it originally appeared in the stimulus tableau (1) or not (0).

To extract and classify errors, we used an automated procedure similar to that of Anderson & Dell (2018); however, since the beats in our fast repetitions were spaced 500ms apart instead of 400ms apart, we used the empirical distribution of keypresses to set the analysis window around each critical beat. The analysis window around a metronome beat contains all keypresses that are aligned to the target key corresponding to that beat. To obtain a reasonable window size, we first extracted the start times of the first and third metronome beats in each pane from the fast repetitions of the sequence—our critical beats, which corresponded to the finger keys. We then calculated the time difference between these metronome beats and each keypress that occurred within 500ms of it, and binned each of these differences into 10ms frames. We summed the distributions across the critical beats to identify the frames with the fewest keypresses, on either side of the metronome beat (see Figure 17). We took these frames as our lower and upper bounds, resulting in an analysis window starting 210ms before each critical beat, and ending 330ms after it.[13]

[13] Compared to Anderson & Dell (2018), this window is shifted slightly to the right.

Figure 17

Distribution of Keypresses and Analysis Window around All Critical Beats



Note. Histogram is summed across all critical beats and participants. Critical beats are only the first and third beat of each pane, within fast repetitions of the sequence. The pink box represents the analysis window that was used to align keypresses to targets for error extraction.

To identify errors, we first extracted all non-target keypresses that occurred within analysis windows around the first and third metronome beat of each pane, during fast repetitions of each stimulus tableau. Since we were interested in errors on finger keys only, we removed thumb keypresses (N = 5,498) and all presses of keys outside the experiment (N = 140) from this sample. We classified each of the remaining errors as position-preserving if the key's original location within the pane where it appeared in the stimulus was the same as that of the beat during which it was pressed. For example, if a key was pressed on the third beat of a pane, but originally appeared as the first key in a pane of the stimulus tableau, the error was classified as non-position-preserving. Conversely, if that key appeared as the third key in a different pane, we classified the error as position-preserving.

To compare our findings to those of the speech experiments (see Chapter 2), we used only errors on keys that originally appeared in the position to which restricted consonants were constrained, for each participant condition. In contrast to our findings in the speech experiments, however (see Chapter 2), coda targets were not more likely to preserve position than onset targets, according to a visual inspection. This procedure resulted in $N = 1,916$ errors available for analysis.

3.4.1.1.2 *Fixed Effects.* Fixed effects were session, condition, and two factors encoding contrasts among finger key targets in the context of each stimulus. We also included interactions between all factors except those encoding contrasts among finger key targets in the context of each stimulus.

Contrast coding for each factor was centered on zero, such that the intercept of the model reflected the grand mean. We defined *session* as a within-participant factor with two effect-coded levels: pre-consolidation (Day 1), coded as -0.5, and post-consolidation (Day 2), coded as 0.5. We defined *restriction position* as a between-participants factor with two effect-coded levels: *restricted-to-last*, coded as 0.5, and *restricted-to-first*, coded as -0.5.

In contrast to previous button-pressing studies that constrained restricted finger key targets to a certain position in every trial, our study only constrained restricted finger key targets to a certain position in the presence of their respective constraining thumb key targets. Our statistical model therefore took into account differences in context among restricted finger key targets by using a two-factor Helmert coding scheme (see Table 11), with separate fixed effects for the difference between restricted and unrestricted finger key targets (*unrestricted vs.*

restricted) and the difference between restricted finger key targets whose position was constrained in the current vowel context and restricted finger key targets whose position was not restricted in the current trial (*restricted-constrained vs. restricted-free*).

Table 11

Contrast Coding Scheme for Finger Key Target Type and Context

Finger key target type	Constraining context	Number of finger key targets per trial	Fixed effect coding	
			unrestricted vs. restricted	*restricted-constrained vs. restricted-free*
experimentally-restricted	constrained = in the context of its thumb key target	1	-0.66	0.5
	unconstrained = in the context of the other (non-conditioning) thumb key target	1	-0.66	-0.5
unrestricted	N/A	4	0.33	0

Note. Predictors used in the model to encode consonant type and context. Each fixed effect was coded such that the sum of values of across consonants was zero.

3.4.1.1.3 *Random Effects Structure, Convergence, and Parsimony.* In addition to aforementioned fixed effects, we initially included random intercepts and slopes for all fixed effects by participant.[14] We used the BOBYQA optimizer (Powell, 2009) to fit model parameters and perform convergence checks.

Because the maximal model did not converge within 100,000 iterations, we constrained correlation parameters to zero, following the recommendations in Bates et al. (2015) to avoid overparameterization.[15] Due to singular fit issues, we removed random slopes by participant for all second- and third-order interactions, as well as the random slope for *restricted-constrained vs. restricted-free*.

3.4.1.1.4 *Significance Testing.* The significance of each predictor was assessed via a likelihood ratio test, in which we compared the likelihood of the data under the full model to the likelihood of the data under a model that did not include the predictor in question via a χ^2 test.

3.4.1.2 Fitted Model Results. On average, approximately 41.94% of keypresses produced in errors preserved their syllable positions (β = -0.56, SE = 0.12). We found a significant main effect of *restricted-constrained vs. restricted-free*, as well as significant two-way interactions between *session* and each of the first-order predictors encoding contrasts among finger key targets in the context of each stimulus. In addition, the three-way interaction between *restricted-constrained vs. restricted-free*, *restriction position*, and *session* was significant. Models that dropped each aforementioned factor (yet retained all other factors) resulted in significant reductions in fit to the data compared to the full model, according to chi-squared tests (see Table 12 for chi-squared test results and coefficient estimates from the statistical model).

---

[14] In our pre-registration, we initially stated that we would use the maximal random effects structure. However, due to the relatively homogenous pool from which stimulus items were chosen, as well as the comparatively small number of observations per stimulus tableau (*M* = 2.29, *SD* = 2.10), relative to the number of observations per participant (*M* = 82.94, *SD* = 134.68), we included only by-participant random effects in the model, opting to forgo by-stimulus random effects.

[15] We followed the recommendations in (Bates et al., 2015) with a single exception: we did not further reduce the random effects structure beyond the number of principal components that cumulatively accounted for 100% of the variance in random effects. This is the most conservative version of the model.

Table 12

Button-Pressing Error Mixed-Effects Model Results

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	-0.56	0.12		
session	-0.22	0.2	1.18	0.276
restricted-constrained vs. restricted-free	0.35	0.15	5.71	0.017 *
unrestricted vs. restricted	-0.19	0.2	0.87	0.35
restriction position	-0.23	0.25	0.84	0.359
restricted-constrained vs. restricted-free × session	-0.69	0.29	5.57	0.018 *
restricted-constrained vs. restricted-free × restriction position	-0.23	0.29	0.59	0.442
unrestricted vs. restricted × session	-0.49	0.24	4.21	0.040 *
unrestricted vs. restricted × restriction position	0.04	0.39	0.01	0.911
restriction position × session	0.04	0.41	0.01	0.916
restricted-constrained vs. restricted-free × restriction position × session	1.29	0.58	4.90	0.027 *
unrestricted vs. restricted × restriction position × session	-0.09	0.48	0.04	0.846

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 3.4.1.1.4.

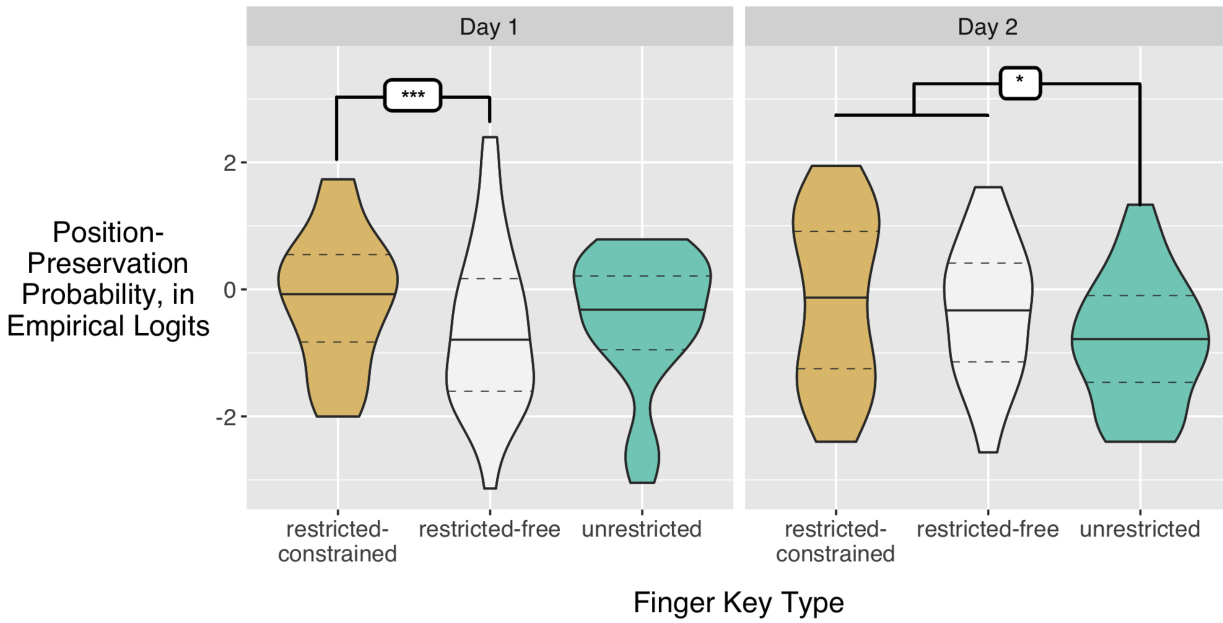
Restricted finger key targets were 1.51 times more likely ($\beta = 0.35$, SE = 0.15) to preserve position when they were in their constrained context, compared to their unconstrained context, indicating that participants did learn to associate restricted finger key targets in the context of their constraining thumb key targets. Although we observed no significant main effect of *unrestricted vs. restricted*, we did find a significant interaction of this factor with *session* (β = -0.49, SE = 0.24), suggesting that participants may have only made this distinction on a single day.

Since we observed significant interactions between *session* and three other predictors (*unrestricted vs. restricted; restricted-constrained vs. restricted-free;* and the *restricted-constrained vs. restricted-free* × *restriction position* interaction), we ran follow-up mixed-effects logistic regressions that included all factors except *session* and its interactions on data from each day.[16] The *restricted-constrained vs. restricted-free* factor was only significant in the Day 1 model ($\beta = 0.35$, SE = 0.15; $\chi^2(1) = 16.05$, $p < 0.001$), while the main effect of *unrestricted vs. restricted* was only significant on Day 2 ($\beta = -0.54$, SE = 0.28; $\chi^2(1) = 3.87$, $p < 0.05$), with unrestricted finger key targets maintaining position only 0.58 times as often as restricted finger key targets (see Figure 18). From Day 1 to Day 2, the position-preservation probability of restricted-free finger key targets increased to match that of restricted-constrained finger key targets, suggesting that the consolidation period may have caused participants to generalize restrictions from constrained contexts to unconstrained (free) contexts.

[16] We followed all procedures for ensuring convergence, avoiding overparameterization, and calculating significance as in the main model. As a result, we constrained correlation parameters to zero in both models, and reduced the random effects structure to exclude all second-order random slopes and the by-participant slope of *restricted-constrained vs. restricted-free* in both models. The Day 1 model was trained on $N = 1509$ observations, and the Day 2 model was trained on $N = 1145$ observations.

Figure 18

Position-Preservation Probability of Finger Key Targets, by Day and Restrictedness



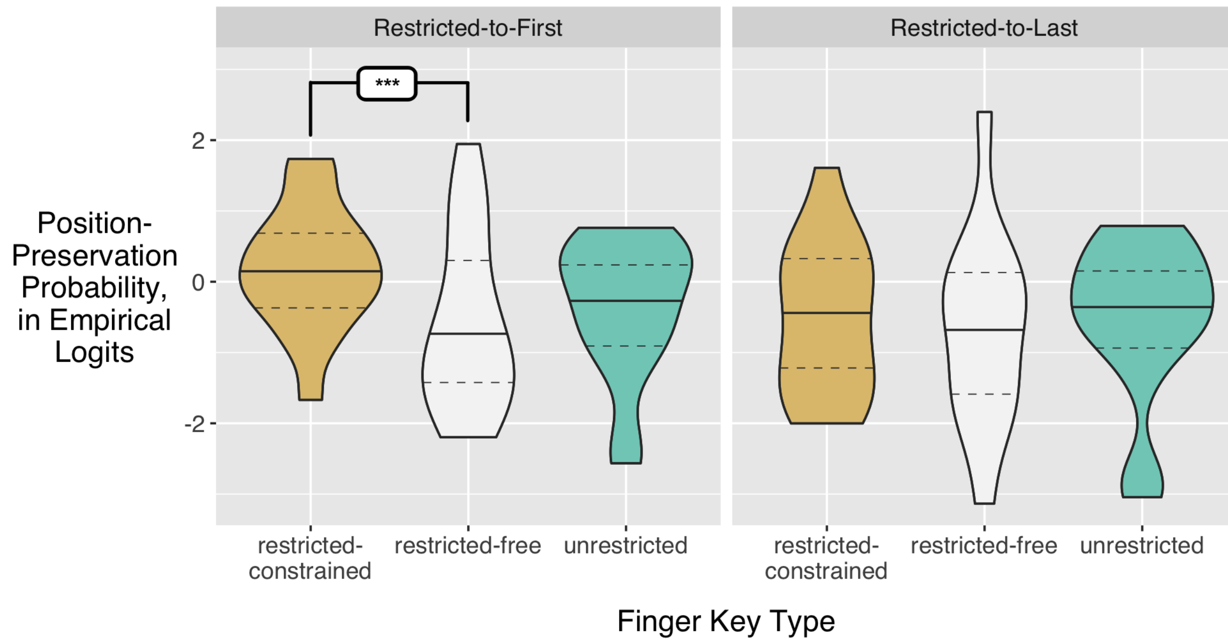
Note. For clarity, only within-day effects are shown. Position-preservation probability was calculated within each session. The area of each violin represents the total number of sessions. Horizontal lines show quantiles.

***$p < 0.001$, *$p < 0.05$

In addition, the Day 1 model revealed a significant interaction between *restricted-constrained vs. restricted-free* and *restricted position* (see Figure 19), with participants in the *restricted-to-last* condition maintaining position on restricted finger key targets in constrained contexts only 0.42 times as often as expected in the absence of this interaction (β = -0.86, SE = 0.36; $\chi^2(1) = 5.87$, $p < 0.05$). To further investigate this interaction, we split the Day 1 data according to participant condition (and again reduced the model formula, this time excluding *session* and its interactions). The *restricted-free vs. restricted-constrained* main effect was only significant in the model fit to errors from participants in the *restricted-to-first* condition, who were approximately 3.14 times more likely to preserve position on errors involving restricted key targets in their constrained position, compared to when those restricted key targets appeared in their unconstrained position (β = 1.14, SE = 0.25; $\chi^2(1)$ = 22.30, $p <$ 0.001). By contrast, participants in the *restricted-to-last* condition were not significantly more likely to preserve position on errors involving restricted key targets in their constrained position (β = 0.24, SE = 0.25; $\chi^2(1)$ = 0.91, p = 0.341). These outcomes indicate that restrictions which bind finger keys to first position are more readily learned, suggesting that the advantage in position-preservation of onset consonants in the tongue-twister task (see Chapter 2) is language-specific. The fact that we did not observe this difference between conditions on Day 2 (despite comparable power) suggests that consolidation may help participants in the restricted-to-last condition "catch up" such that the restrictions could inform their button-presses on Day 2—though this may also be an artifact of the greater number of errors made on that day by errors in that condition (see Section 3.4.1.3 below).

Figure 19

Position-Preservation Probability of Restricted Finger Key Targets on Day 1, by Condition



Note. Data from Day 1 only. For clarity, only within-condition effects are shown. Position-preservation probability was calculated for each participant's Day 1 session. The area of each violin represents the total number of sessions in which at least one error on a finger key target of the type shown on the x-axis was made. Horizontal lines show quantiles.

$^{***}p < 0.001$

3.4.1.3 Number of Errors. On average, participants produced 86.36 errors per session (minimum = 5; maximum = 1376). In order to test whether the number of errors in each session may have mediated the effects we found in our analysis, we first tested whether the number of errors per session correlated with either *restriction position* or *session* by fitting a mixed-effects negative binomial linear model that predicted number of errors from *restriction position*, *session*, and their interaction. Using the model simplification procedures described in Bates et al. (2015), we were able to retain random intercepts by participant in the model. We used the same significance testing procedure described in Section 2.3.2.2.4 to test whether each predictor was significant. Results are shown in Table 13.

Table 13

Button-Pressing Experiment Number of Errors Model Results

Predictor	Estimate	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	4.35	0.14		
session	-0.50	0.27	3.24	0.072
restriction position	-0.41	0.27	2.28	0.131
restriction position × *session*	-1.37	0.54	0.12	< 0.05 *

Note. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4.

We found a significant interaction between of *session* and *restriction position*, such that participants in the restricted-to-last condition made significantly fewer errors on Day 2 than would be expected, in the absence of this interaction. To investigate the extent to which this may have influenced the outcome of our main model, we ran a follow-up version of the model that included all predictors from the original model, plus the main effect of *number of errors* (transformed via the natural log, then centered on a mean of 0 and scaled to a standard deviation of 1) and its interaction with *restricted-constrained vs. restricted-free*. The *restricted-constrained* *vs. restricted-free* effect was still significant in this follow-up model (β = 0.34, SE = 0.15; $\chi^2(1)$ = 5.40, p < 0.05), as was its interaction with *session* (β = -0.66, SE = 0.30; $\chi^2(1)$ = 4.86, p < 0.05). The interaction between *session* and *unrestricted vs. restricted* was also significant (β = -0.48, SE = 0.24; $\chi^2(1)$ = 4.08, p < 0.05).

However, the three-way interaction between *restricted-constrained vs. restricted-free*, *session*, and *restriction condition* was not significant (β = 1.12, SE = 0.64; $\chi^2(1)$ = 3.06, p = 0.080) with the addition of the interaction between *restricted-constrained vs. restricted-free* and number of errors. This result suggests that the putative consolidation benefit in the restricted-to-last condition may instead be an artifact of the greater number of errors on Day 2 in that condition, compared to the restricted-to-first condition.

3.4.2 *Sleep Data*

3.4.2.1 Data Quality. We collected biometric sleep data from 13 participants between Day 1 and Day 2 using activity trackers. Of these, 11 participants (84.62%) provided their height, weight, age, and gender when signing up for the study. Since Fitbit uses this demographic information to calculate heart rate zones for each participant, (M. Considine, personal communication, March 19, 2021), we omitted the 2 participants who did not provide full demographic information from our analysis of sleep data.

3.4.2.2 Testing for Consolidation Effects. To attempt to establish whether this effect is tied to consolidation, we constructed another logistic regression predicting position-preservation of errors in which we replaced the *session* predictor from the original model with the interaction between *session* and the z-score corresponding to the proportion of time spent in slow-wave sleep between sessions (*SWS proportion*). Since only 11 participants whose data was included in this experiment also recorded biometric sleep data and provided full demographic information, this comparison was performed only on data from those 11 participants. We then compared the fit of this model to that of the original model (described in Section 3.4.1.1) on the same data, using the likelihood ratio test described in Section 3.4.1.1.4. The addition of slow-wave sleep as a predictor did not significantly improve fit, according to a likelihood ratio test between the slow-wave sleep model and the full model ($\chi^2(1) = 3.85$, $p = 0.427$), so we cannot conclude that this difference between days is due to slow-wave sleep. That being said, the interactions with *session* were also not significant in the baseline (original) model for this 11-participant sample, so it is possible that this follow-up analysis was hampered by a lack of statistical power.

3.5 Discussion

This experiment investigated the possibility that effects found in Experiment 1 of Chapter 2 may have been due to domain-general properties of sequential pattern learning. Table 14 summarizes our findings.

Table 14

Predictions for Button-Pressing Experiment (Repeated from Table 9) and Outcomes

Result from speech experiments	Proposed explanation(s) for speech experiment result	Prediction (if phonotactic learning results are solely driven by sequential pattern learning)	Evidence for prediction found?
In contrast to previous studies, participants were able to restrict positions of constrained consonants on Day 1.	Lack of interference between consonant-position mappings.	Similar; participants should be able to restrict the position of finger keys targeted by the restriction on Day 1.	Yes
Experimentally-restricted consonants preserved position more often than unrestricted consonants, even in in vowel contexts that did not constrain their positions.	Mappings between response units and their positions are represented both locally (with adjacent context) and globally (independent of adjacent response units).	Similar; restricted finger key targets in unrestricted (free) contexts will preserve position more often than unrestricted consonants.	Yes, but only on Day 2
Participants preserved position more often on coda consonant errors than onset consonant errors.	Planning differences between the first and third response units.	Similar; third-key errors will preserve position more often than first-key errors.	No

3.5.1 *Consolidation Resolves Interference among Conflicting Mappings Between Response Units and Sequential Position*

On Day 1, participants' errors on restricted finger key targets in constrained trial contexts preserved position significantly more often than their errors on the same consonants in unconstrained trial contexts. This indicated that participants did successfully learn the phonotactic restrictions on Day 1, suggesting that the lack of conflict among mappings between response units and sequential positions may promote sequential pattern learning more generally. The lack of improvement in this metric after consolidation implies that consolidation may have resolved the mapping conflict in Anderson and Dell (2018) and Rebei et al. (2019), in contrast to this experiment. Nevertheless, further research using a controlled experiment and a common statistical model is necessary to confirm this, since the lack of improvement can stem from many sources.

3.5.2 *Consolidation Aids in Schema Creation, Promotes Transfer across Trial Contexts*

In this experiment, restricted finger key targets in unrestricted (free) contexts will preserve position more often than unrestricted consonants, but only after an opportunity for sleep-based consolidation. We take this as evidence that the opportunity for sleep-based consolidation creates novel, structured representations of targets that participants can generalize across to improve performance—i.e., new schema. In contrast to the speech experiments (Chapter 2), participants bring very little (if any) prior knowledge to this experiment. Therefore, participants must learn to chunk each pane in the sequence into a hierarchical representation in order to efficiently represent the panes (Sakai et al., 2003) and achieve optimal performance on the task. Sleep-based consolidation appears to aid in the development of this hierarchical schema, as expected under Stickgold and Walker (2013)'s selective processing account of consolidation. This is compatible with previous research on schema development that implicates sleep-dependent consolidation, as well as the brain structures typically active during consolidation (Tse et al., 2007).

What evidence do we have to support schema creation? On Day 1, the primary distinction that participants make is among restricted finger key targets. Errors containing restricted finger key targets preserve position more often when they appear in constrained thumb contexts, compared to when they appear in unconstrained thumb contexts. There was little difference between unrestricted consonants and restricted consonants in unconstrained contexts, signaling that participants' errors faithfully mirrored the distribution of targets, according to the second-order regularities present in the stimuli. On Day 2, however, the *unrestricted vs. restricted* distinction was in play: participants' errors preserved position more often on restricted finger key targets, compared to unrestricted finger key targets. This separation by finger key class suggests that consolidation has allowed participants to integrate statistical information regarding position of appearance in order to create an efficient representation in which the two classes of target response units (restricted and unrestricted) have different distributional properties. Moreover, the distinction among restricted consonants based on context is attenuated: transfer across contexts within finger key targets has occurred, as we would expect if participants had learned to chunk and associate each finger key with the position it occurs in most often, globally.

3.5.3 *Tendency to Keep Rhymes Intact in Errors May Be Language-Specific*

We did not replicate the difference by participant condition from the speech experiments; there was no significant main effect of condition on position-preservation in errors, and specifically, no evidence of greater position-preservation in errors made by participants whose restricted finger key targets were restricted to the final position, within each pane. If anything, participants in the *restricted-to-first* condition were more likely to preserve position on restricted key targets in their constrained position than those in the *restricted-to-last* condition. Although our dataset contained fewer errors in this experiment (and therefore less power), we did find significant effects in this experiment that were smaller than the effects of condition in the speech experiments, leading us to believe that if this data did contain a difference of comparable size, we would be able to detect it using our statistical method. Ultimately, this means that the tendency to keep syllable rhymes intact in errors may be an English-specific property due to the high predictability of codas, given onsets (Kessler & Treiman, 1997), though future research should attempt to falsify this using participants from different native language backgrounds.

3.5.4 *Methodological Innovation in Remote Research*

On a practical note, this experiment is the first button-pressing regularity study that uses home computers and adapted QWERTY keyboards for data collection. We present this as a proof-of-concept for researchers who may be facing barriers to lab-based data collection using button-boxes and sound-isolated booths—be it from pandemics, budgetary concerns, or other obstacles. Future studies should attempt to replicate well-attested lab-based findings remotely, in the vein of recent remote replications of speech production work (see Fairs & Strijkers, 2021 for an example). This work has the potential to allow researchers to collect data from understudied populations of participants for whom lab-based data collection is not feasible, thereby capturing a more complete picture of human cognitive diversity and empowering more individuals to participate in scientific advancement.

3.6 Conclusion

In this experiment, we tested whether the functions of consolidation that we identified in our speech experiments were also operative in a domain with little (if any) prior structured knowledge: a sequential button-pressing task. We found evidence of another role for consolidation: schema creation, after which transfer of distributional information among identical response units across contexts is possible. We do not find evidence of the general tendency to keep rhymes (second + third sequential positions) together in errors, which was present in our earlier speech experiments (Chapter 2); this suggests that the tendency is a linguistically-driven effect, possibly arising from distributional characteristics of participants' native language. Overall, in the absence of prior knowledge regarding sequential regularities, sleep-based consolidation allows participants to chunk information into a representation that is helpful for future production.

Conclusion

In this work, we examined the role of prior knowledge in the consolidation of second-order phonotactic restrictions for speech production. We tested the hypothesis that new phonotactic restrictions which are dissimilar to previously-learned phonotactic generalizations may be more difficult to consolidate during sleep, and therefore benefit less from sleep-based consolidation. We investigated this in a speech production task (Chapter 2) as well as in a non-linguistic sequential action task (Chapter 3) where participants had no structured prior knowledge, in order to assess whether results from our speech production task were due to properties of the linguistic system, or to more domain-general planning and production processes.

In Chapter 2, we tested two sources of prior phonotactic knowledge: participants' native language (Experiment 1) and pre-training in the experimental setting (Experiment 2). We found that consolidation did enable transfer from prior phonotactic knowledge, but only when that knowledge was acquired in a similar environment as the task itself (i.e., pre-training). In contrast to previous second-order phonotactic learning studies where mappings between restricted consonants and the positions they appeared in interfered with each other, we found evidence of phonotactic learning on Day 1 of each experiment, suggesting that consolidation's role in previous studies had been to resolve interference between conflicting mappings. Participants tended to keep syllable rimes intact in speech errors; we failed to replicate this in a non-linguistic sequential action task (button-pressing; see Chapter 3), suggesting that this tendency stemmed from distributional characteristics of participants' native language (English). We additionally found evidence that participants generalized second-order phonotactic restrictions to context-independent first-order constraints that bound restricted consonants to a particular syllable position more often than chance even when not constrained by their immediate vowel context; in button-pressing, this generalization was only evident after a consolidation period.

4.1 Implications

4.1.1 *Representing Phonotactic Patterns*

Across both our speech experiments, we found that participants tended to produce restricted consonants in their restricted positions, even in stimuli where their position was not constrained by the vowel. This indicates that participants represent associations between consonants and syllable positions at two levels: global, context-independent, first-order generalizations (*e.g.*, "/f/ appears in coda position 75% of the time") and local, context-dependent, second-order phonotactic generalizations (*e.g.*, "/f/ is more likely to appear in coda position if /ɪ/ is the vowel). These representations overlap to affect production; in our example, /f/ would be most likely to appear in coda in errors on stimuli containing /ɪ/, but also more likely to appear in coda in errors on stimuli containing /æ/ than an unrestricted consonant. This is the case even after an opportunity for sleep-based consolidation, indicating that consolidation may not be able to perfectly separate these contexts to create a veridical representation of consonant distributions. This effect may arise from the fact that both generalizations are largely consistent with the presented data, and so both are simultaneously used in production.

We found transfer effects in our second speech experiment, such that the difference in position-preservation rate on restricted consonants by vowel context was larger for participants in the *training-similar* group than for those in the *training-dissimilar* group. The similarity among restrictions across phases gave the *training-similar* group an advantage in learning. Results from the *training-dissimilar* group, on the other hand, suggested that this group needed to unlearn previously-formed associations between restricted consonants and the syllable positions to which they had been constrained in pre-training before new, conflicting restrictions could affect speech production. These conclusions, along with Anderson et al. (2019)'s findings regarding reversal shift in phonotactic learning, support an incremental account of phonotactic learning.

4.1.2 *Role of Sleep-Based Consolidation Depends on Task and Extant Prior Knowledge*

Each of our experiments uncovered a different role for consolidation, depending on the characteristics of the task, as well as the nature of prior knowledge. In our first speech experiment, the role of consolidation was minimal, despite our large sample size; participants' speech errors followed the phonotactic restrictions in question even on Day 1. This was surprising, since previous studies of second-order phonotactic learning had found consolidation effects (Gaskell et al., 2014; Smalle et al., 2017, 2021; Warker, 2013; Warker et al., 2010; Warker & Dell, 2006) in undistracted adults. We believe this is due to a difference in how our stimuli were constructed. In previous experiments, the mapping between each restricted consonant and its syllable position was completely reversed on trials containing a different vowel; therefore, participants were tasked with learning two conflicting associations between consonant and syllable position. In our stimuli, by contrast, both consonants could appear in the targeted position in any trial, but each consonant was additionally permitted to appear in the opposite position on trials that did not include its constraining vowel. The mapping between unconstrained restricted consonants and position targeted by the restriction only needed to be inhibited on 25% for successful (errorless) performance. This lessened inhibition on Day 1 may have reduced the need for participants to foreground previously-backgrounded information during consolidation, as they would under situations of direct conflict (see Smalle et al., 2021 for an example and discussion of executive control's role in backgrounding). This account explains why participants could successfully learn the phonotactic restrictions in this experiment before consolidation. In light of previous work, the results of this experiment suggest that consolidation's role in past studies has been to resolve interference between directly conflicting consonant-position mappings.

In our second speech experiment, consolidation made previously-learned information available for speech production. Specifically, the difference in position-preservation rate between errors on restricted and unrestricted consonants was larger after an opportunity for sleep-based consolidation; in addition, the increase in position-preservation rate between errors on restricted consonants in constrained context compared to unconstrained context was larger for participants in the training-dissimilar condition after an opportunity for sleep-based consolidation. The difference between this experiment and Experiment 1 suggests that consolidation does enable participants to draw upon prior knowledge for speech production, but only when that prior knowledge has been acquired or activated[17] in a similar environmental context. This supports a view of phonotactic learning in which restrictions are stored in tandem with their contexts of acquisition. We suspect that effect was not observed in Experiment 1 because participants' native language was acquired and mainly used in naturalistic environments that were very different from the experimental context. Taken together, these results support Dell et al. (2019)'s proposal that sleep-based consolidation makes a contextually-bound mini-grammar available for speech production.

[17] Our paradigm did not disambiguate the context of acquisition from the context of usage/activation. Thus, the extent to which the initial learning environment may be privileged with respect to phonotactic transfer remains to be seen. It is possible that this effect is driven by similarity between usage environments instead.

Our button-pressing results suggest that consolidation may have created a new schema, such that responses were more in line with our speech production results after a period of consolidation. On Day 1 of the button-pressing experiment, participants had not generalized the restriction applicable to restricted finger key targets in constrained contexts to the same targets in different trial contexts. This suggests that consolidation allowed participants to integrate statistical information regarding position of appearance in order to create an efficient representation that represents both first- and second-order sequential generalizations simultaneously, as in the speech experiments. Together with results from the speech experiments, our evidence suggests that consolidation was needed because participants lacked the structured prior knowledge from a lifetime producing syllables; consolidation thus created a "syllable-like" schema for button-presses.

4.1.3 ***Possible Effects of Native Language on Speech Errors***

In both of our speech experiments, participants' errors on coda consonants tended to preserve position more often than their errors on onset consonants. This effect was not present on either day of the button-pressing experiment, despite comparable power, indicating that it does not arise due to domain-general sequential action planning processes. Instead, we propose that it may be due to a general property of English: since codas are more predictable from vowels than onsets (Kessler & Treiman, 1997), there may be a tendency to keep the rime unit intact in speech errors. Studies of second-order phonotactic learning in adults have found differences by native language (Smalle & Szmalec, 2021), so it is possible that the task may reflect other characteristics of speakers' native languages—especially since the vast majority of studies using this paradigm have only been conducted on native English speakers.

4.2 Limitations

Although we took steps to minimize the amount of noise present in our speech data, agreement on errors was only 55.54% (Cohen's K = 0.49) across both experiments. Recording in a cleaner environment (e.g., a sound-isolated recording booth with a high-quality microphone) may have yielded more consistent data than our remote data collection procedure.[18] Nevertheless, since statistics on errors are not usually reported separately from agreement metrics on non-errors (and our agreement on non-errors was almost perfect), we cannot at present establish the amount of noise that remote recording may have introduced. As a step toward minimizing noise and potentially quantifying the limitations of the tongue-twister paradigm, we recommend that other researchers report agreement statistics on errors alone and use a blind transcription procedure and a common error-detection algorithm (such as the one described in Section 2.3.2.2.1) to minimize the effects of coder bias and human error.

Our studies also suffered from a lack of power in biometric sleep data. Due to logistical concerns regarding synchronizing the activity monitors during drop-off, participants were not compensated separately for providing sleep data; instead, their compensation solely depended on completion of the experimental sessions. We recommend that future studies employing a similar hybrid data collection method include a bonus payment in their compensation scheme in order to incentivize participants to wear the activity trackers between sessions.

[18] Goldrick (2004) found an agreement rate of 77.8% on errors from a tongue-twister task with a similar consonant inventory that was recorded in a sound-attenuated chamber. Comparison to our agreement rate suggests that our remote recording procedure did introduce noise. However, Goldrick (2004)'s dataset only contained 9 jointly transcribed errors (compared to 3001 across both our speech experiments). More data would provide a less noisy agreement estimate.

4.3 Future Directions

Results from our second speech experiment suggest that similarity between learning environments can encourage integration of information into an existing schema during consolidation. This raises several questions regarding the dimensions of "similarity" that are relevant for integration during consolidation. How similar must two environmental contexts be for integration? Under a selective processing account of consolidation (Stickgold & Walker, 2013), task-relevant similarities (e.g., the consonant inventory in tongue-twisters) should be more critical for promoting integration during consolidation than task-irrelevant similarities (e.g., the color of the text on the screen). Future research should examine this possibility, in order to determine the extent to which contextual information can affect consolidation. Beyond its intellectual significance, this research could have practical implications for language pedagogy.

Dell et al. (2019)'s theory implies that participants created an initial experimental mini-grammar during the first speech experiment that could be "siloed off" from the rest of the English language, such that updates to that mini-grammar did not change participants' English phonotactics. Since the circumstances that trigger formation of a new mini-grammar are undefined, it is unclear whether this was the case for the second speech experiment as well. Was the replacement of two consonants in the inventory a sufficiently salient context cue to trigger the creation of new, contextually-bound mini-grammar for Experiment 2, separate from the Experiment 1 mini-grammar? If this is the case, then the restrictions learned in Experiment 2 should not override those from Experiment 1; if participants were again exposed to tongue-twisters with the same consonant inventory and phonotactic restrictions as in Experiment 1, they should have little difficulty reverting to their Experiment 1 mini-grammar, even if they were exposed to conflicting restrictions in Experiment 2.

Defining the precise neurocognitive mechanism that facilitated this integration of information from pre-training may help clarify this point, as well as other open questions regarding the role of prior knowledge in phonotactic learning for speech production. For example, neuronal replay during consolidation (Lewis & Durrant, 2011) may help participants identify overlapping unrestricted consonants between the two phases of the experiment, thereby permitting the insight that these consonants have the same distribution, regardless of the new consonants in the tongue-twister. Consolidation may also help participants link shared phonological properties (in this case, manner and place of articulation) between restricted consonants in the two phases of the experiment. These possibilities are not exclusive; future research should test them separately. Discovering the source of these effects would clarify the role of consolidation in language learning, with potential implications for learning of other linguistic phenomena as well.

Finally, the reason for simultaneous representation of first- and second-order generalizations throughout our experiment is still unclear. Results from the button-pressing experiment suggest that this pattern is an artifact of the production schema, which in turn may mean that the generalization within identical response units may be necessary to efficiently represent second-order sequential patterns, even beyond speech production. It remains to be seen whether this generalization is only useful during the early stages of acquisition (and thus can be overcome with sufficient exposure), or whether it arises from other biases that persist as long as the schema for production is in place. Clarifying this point can have important implications for the study of sequential action patterns and how they are learned.

4.4 Conclusion

This work investigated prior knowledge's role in consolidation of second-order phonotactics for speech production, by analyzing production errors made in two sequential action domains. Our results indicate that first-order and second-order sequential patterns are simultaneously represented in production, and that this may be a necessary property of the schema for structured sequential action, in both speech and keypress sequence production. We found evidence of transfer from phonotactic restrictions that had been previously learned in a similar environmental context, supporting an incremental account of phonotactic learning (Anderson et al., 2019) as well as a view of phonotactic learning in which restrictions are stored together with their contexts of acquisition (Dell et al., 2019). Our results also suggest that consolidation's main function in phonotactic learning for speech production is to resolve interference among multiple conflicting mappings between consonants and their syllable positions. Moreover, we found evidence that participants' native language (English) may have affected their speech errors. More research is necessary to clarify the function of environmental context in phonotactic transfer and learning, as well as the neural mechanism underpinning phonotactic transfer and the timecourse of generalization (and possible eventual inhibition) across contexts in non-self-interfering second-order sequential patterns. Overall, this research demonstrates that prior knowledge's complex role in sequential pattern learning, which depends not only on similarity to new information and opportunity for consolidation, but also on the environment in which that prior knowledge was acquired.

References

Anderson, N. D., & Dell, G. S. (2018). The role of consolidation in learning context-dependent phonotactic patterns in speech and digital sequence production. *Proceedings of the National Academy of Sciences*, *115*(14), 3617–3622. https://doi.org/10.1073/pnas.1721107115

Anderson, N. D., Holmes, E. W., Dell, G. S., & Middleton, E. L. (2019). Reversal shift in phonotactic learning during language production: Evidence for incremental learning. *Journal of Memory and Language*, *106*, 135–149. https://doi.org/10.1016/j.jml.2019.03.002

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *ArXiv:1506.04967 [Stat]*. http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boomer, D. S., & Laver, J. D. M. (1968). Slips of the Tongue. *International Journal of Language & Communication Disorders*, *3*(1), 2–12. https://doi.org/10.3109/13682826809011435

Carlisle, R. S. (1988). The Effect of Markedness on Epenthesis in Spanish/English Interlanguage Phonology. *Issues and Developments in English and Applied Linguistics (IDEAL)*, *3*, 15–23.

Carlisle, R. S., & Cutillas Espinoza, J. A. (2014). The Production of /.sC/ Onsets in a

Markedness Relationship: Investigating the Ontogeny Phylogeny Model with Longitudinal Data. In Mehmet Yavas (Ed.), *Unusual Productions in Phonology: Universals and Language-Specific Considerations*. Psychology Press.

Cedrus Response Pad RB-840. (n.d.). NEUROSPEC Webshop. Retrieved September 29, 2022, from https://shop.neurospec.com/cedrus-response-pad-rb-840

Considine, M. (2021, March 19). *RE: Fitbits for Research* [Personal communication].

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*(2), 149–195. https://doi.org/10.1016/0364-0213(93)90010-6

Dell, G. S., Kelley, A. C., Bian, Y., & Holmes, E. W. (2019). Tuning the blueprint: How studies of implicit learning during speaking reveal the information processing components of the production system. *Language, Cognition and Neuroscience*, *34*(9), 1246–1256. https://doi.org/10.1080/23273798.2019.1613553

Dell, G. S., Kelley, A. C., Hwang, S., & Bian, Y. (2021). The adaptable speaker: A theory of implicit learning in language production. *Psychological Review*. https://doi.org/10.1037/rev0000275

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(6), 1355–1367. https://doi.org/10.1037//0278-7393.26.6.1355

Diekelmann, S., & Born, J. (2010). Slow-wave sleep takes the leading role in memory reorganization. *Nature Reviews Neuroscience*, *11*(3), Article 3. https://doi.org/10.1038/nrn2762-c2

Fairs, A., & Strijkers, K. (2021). Can we use the internet to study speech production? Yes we can! Evidence contrasting online versus laboratory naming latencies and errors. *PLOS ONE*, *16*(10), e0258908. https://doi.org/10.1371/journal.pone.0258908

Feinerer, I., & Hornik, K. (2020). *wordnet: WordNet Interface* (R package version 0.1-15). https://CRAN.R-project.org/package=wordnet

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.

Fitbit, Inc. (n.d.). *How do I wear my Fitbit device?* Fitbit Help. Retrieved May 24, 2021, from https://help.fitbit.com/articles/en_US/Help_article/1988.htm

Fromkin, V. A. (1971). The Non-Anomalous Nature of Anomalous Utterances. *Language*, *47*(1), 27.

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., & Stackhouse, A. (2014). Sleep Underpins the Plasticity of Language Production. *Psychological Science*, *25*(7), 1457–1465. https://doi.org/10.1177/0956797614535937

Genzel, L., & Robertson, E. M. (2015). To Replay, Perchance to Consolidate. *PLOS Biology*, *13*(10), e1002285. https://doi.org/10.1371/journal.pbio.1002285

Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, *51*(4), 586–603. https://doi.org/10.1016/j.jml.2004.07.004

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, *107*(3), 1155–1164. https://doi.org/10.1016/j.cognition.2007.11.009

Haghayegh, S., Khoshnevis, S., Smolensky, M. H., Diller, K. R., & Castriotta, R. J. (2020). Performance assessment of new-generation Fitbit technology in deriving sleep parameters and stages. *Chronobiology International*, *37*(1), 47–59.

https://doi.org/10.1080/07420528.2019.1682006

Harris, J. W. (James W. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, Mass. : MIT Press. http://archive.org/details/syllablestructur00harr

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). *Journal of Memory and Language*, *52*(1), 58–70. https://doi.org/10.1016/j.jml.2004.07.006

Hodne, B. (1985). Yet Another Look at Interlanguage Phonology: The Modification of English Syllable Structure by Native Speakers of Polish. *Language Learning*, *35*(3), 405–417. https://doi.org/10.1111/j.1467-1770.1985.tb01084.x

Kessler, B., & Treiman, R. (1997). Syllable Structure and the Distribution of Phonemes in English Syllables. *Journal of Memory and Language*, *37*(3), 295–311. https://doi.org/10.1006/jmla.1997.2522

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Laubstein, A. S. (1987). Syllable Structure: The Speech Error Evidence. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *32*(4), 339–363. https://doi.org/10.1017/S0008413100012445

Lee, Y., & Goldrick, M. (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language*, *59*(2), 155–168. https://doi.org/10.1016/j.jml.2008.03.002

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, *15*(8), 343–351. https://doi.org/10.1016/j.tics.2011.06.004

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights From the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review*, *102*(3), 419–457.

Mirea, N., & Goldrick, M. (2021). *Recruitment of Prior Knowledge during Sleep-Based Consolidation of Phonotactic Patterns for Speech Production*. OSF. https://doi.org/10.17605/OSF.IO/UNTSH

Motley, M. T. (1973). An analysis of spoonerisms as psycholinguistic phenomena. *Speech Monographs*, *40*(1), 66–71. https://doi.org/10.1080/03637757309375781

Motley, M. T., & Baars, B. J. (1975). Encoding Sensitivities to Phonological Markedness and Transitional Probability: Evidence from Spoonerisms. *Human Communication Research*, *1*(4), 353–361. https://doi.org/10.1111/j.1468-2958.1975.tb00284.x

Newport, E. L. (1990). Maturational Constraints on Language Learning. *Cognitive Science*, *14*(1), 11–28. https://doi.org/10.1207/s15516709cog1401_2

Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives* (p. 39). Cambridge University Department of Applied Mathematics and Theoretical Physics.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rebei, A., Anderson, N. D., & Dell, G. S. (2019). Learning the phonotactics of button pushing: Consolidation, retention, and syllable structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000686

Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Experimental Brain Research*, *152*(2), 229–242. https://doi.org/10.1007/s00221-003-1548-8

Smalle, E. H. M., Muylle, M., Duyck, W., & Szmalec, A. (2021). Less is more: Depleting cognitive resources enhances language learning abilities in adults. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001058

Smalle, E. H. M., Muylle, M., Szmalec, A., & Duyck, W. (2017). The different time course of phonotactic constraint learning in children and adults: Evidence from speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1821–1827. https://doi.org/10.1037/xlm0000405

Smalle, E. H. M., & Szmalec, A. (2021). Quick learning of novel vowel-consonant conjunctions within the mature speech production system – a commentary on Dell et al. (2019). *Language, Cognition and Neuroscience*, 1–5. https://doi.org/10.1080/23273798.2021.1995613

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience*, *16*(2), 139–145. https://doi.org/10.1038/nn.3303

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and Memory Consolidation. *Science*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Wallace, M. (2007). *Jawbone Java WordNet API*. https://sites.google.com/site/mfwallace/jawbone

Warker, J. A. (2013). Investigating the retention and time course of phonotactic constraint learning from production experience. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *39*(1), 96–109. https://doi.org/10.1037/a0028648

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 387–398. https://doi.org/10.1037/0278-7393.32.2.387

Warker, J. A., & Dell, G. S. (2015). New phonotactic constraints learned implicitly by producing syllable strings generalize to the production of new syllables. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(6), 1902–1910. https://doi.org/10.1037/xlm0000143

Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2010). *Limits on Learning Phonotactic Constraints from Recent Production Experience*. 15.

Wells, R. (1973). Predicting slips of the tongue. In V. A. Fromkin (Ed.), *Speech Errors as Linguistic Evidence* (pp. 82–87). Mouton. (Original work published 1951)

Westerlund, A., Brandt, L., Harlid, R., Åkerstedt, T., & Lagerros, Y. T. (2014). Using the Karolinska Sleep Questionnaire to identify obstructive sleep apnea syndrome in a sleep clinic population. *The Clinical Respiratory Journal*, *8*(4), 444–454. https://doi.org/10.1111/crj.12095

Westerlund, A., Lagerros, Y. T., Kecklund, G., Axelsson, J., & Åkerstedt, T. (2013). Relationships Between Questionnaire Ratings of Sleep Quality and Polysomnography in Healthy Adults. *Behavioral Sleep Medicine*, *14*(2), 185–199. https://doi.org/10.1080/15402002.2014.974181

Wilshire, C. E. (1999). The “Tongue Twister” Paradigm as a Technique for Studying Phonological Encoding. *Language and Speech*, *42*(1), 57–82.

Appendix A: Note on Empirical Logits

Many of our violin plots visualize position-preservation probability in empirical logit units. In these figures, each point in each violin is a summary of errors involving a single consonant type (usually indicated along the x axis), within a single session. To create this summary (i.e., to calculate the empirical logit), we use the formula $ln\left(\frac{P + 0.5}{T - P + 0.5}\right)$, where ln is the natural log function, P is the number of position-preserving errors in the session involving the specified consonant type, and T is the total number of errors in the session involving the specified consonant type (Barr, 2008). This formula is roughly equivalent to the standard logit transformation $ln\left(\frac{\frac{P}{T}}{1 - \frac{P}{T}}\right)$, but avoids complications when $\frac{P}{T}$ approaches 0 or 1, since $ln(\frac{0}{1}) = -\infty$ and $\frac{1}{0}$ is undefined. As in the standard logit transformation, an empirical logit value of 0 represents an equal number of position-preserving and position-altering errors ($\frac{P}{T} = 0.5$), and higher values represent a higher proportion of position-preserving errors.

Although we use empirical logits to produce visualizations of our data, we used logistic regression on the raw observations to calculate the significance of our predictors. In other words, empirical logits are used throughout this work to plot position-preservation probability only, and not for statistical analysis.

Appendix B: Free-Context-as-Baseline Model

In order to directly compare position-preservation probabilities of unrestricted consonants to those of restricted consonants in unconstrained (free) vowel contexts, we ran a version of our model on the same data in which the Helmert codings corresponding to consonant type and context were replaced with a 3-level simple contrast coding scheme (see Table B1).

Table B1

Contrast Coding Scheme for Consonant Type and Context, in Free-Context-as-Baseline Model

Consonant type	Constraining context	Number of consonants per trial	Fixed effect coding	
			unrestricted vs. restricted-free	*restricted-constrained vs. restricted-free*
experimentally-restricted	constrained = in the context of its conditioning vowel	1	$-\frac{2}{3}$	$\frac{5}{6}$
	unconstrained = in the context of the other (non-conditioning) vowel	1	$-\frac{2}{3}$	$-\frac{1}{6}$
unrestricted	N/A	4	$\frac{1}{3}$	$-\frac{1}{6}$

Note. Predictors used in the free-context-as-baseline model to encode consonant type and context. Each fixed effect was coded such that the sum of values of across consonants was zero.

Speech Experiments

Experiment 1

Due to convergence issues, we constrained all correlation parameters to zero and removed the following random slopes by participant from the model:

- *restricted-constrained vs. restricted-free* × *session*
- *restricted-constrained vs. restricted-free* × *session* × *similarity to English*
- *unrestricted vs. restricted-free* × *session* × *similarity to English*

Results from the model are shown in Table B2.

Table B2

Experiment 1 Free-Context-As-Baseline Results

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	0.73	0.12		
session	0.02	0.13	0.02	0.889
similarity to English	1.40	0.25	28.00	< 0.001 ***
restricted-constrained vs. restricted-free	1.11	0.25	19.46	< 0.001 ***
unrestricted vs. restricted-free	-1.27	0.21	32.53	< 0.001 ***
similarity to English × *session*	-0.15	0.26	0.29	0.590
restricted-constrained vs. restricted-free × *session*	-0.27	0.41	0.43	0.514
restricted-constrained vs. restricted-free × *similarity to English*	-0.34	0.48	0.39	0.530
unrestricted vs. restricted-free × *session*	-0.36	0.39	0.82	0.365
unrestricted vs. restricted-free × *similarity to English*	-0.48	0.43	1.29	0.256
restricted-constrained vs. restricted-free × *similarity to English* × *session*	-0.78	0.82	0.87	0.352
unrestricted vs. restricted-free × *similarity to English* × *session*	-0.23	0.79	0.08	0.775

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4; p values are not adjusted for multiple comparisons. Predictors were coded according to the scheme in Table B1.

Experiment 2

Due to convergence issues, we constrained all correlation parameters to zero. We were able to retain all random slopes in the model. Results from the model are shown in Table B3.

Table B3

Experiment 2 Free-Context-As-Baseline Results

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	0.80	0.13	NA	NA
session	0.01	0.16	0.01	0.935
similarity to English	-0.96	0.26	12.38	< 0.001 ***
restricted-constrained vs. restricted-free	1.01	0.28	11.35	< 0.001 ***
unrestricted vs. restricted-free	-0.66	0.25	6.33	< 0.05 *
similarity to English × *session*	0.10	0.32	0.09	0.767
restricted-constrained vs. restricted-free × *session*	0.21	0.46	0.21	0.645
restricted-constrained vs. restricted-free × *similarity to English*	1.78	0.56	9.16	< 0.01 **
unrestricted vs. restricted-free × *session*	-0.50	0.42	1.38	0.241
unrestricted vs. restricted-free × *similarity to English*	0.18	0.51	0.13	0.721
restricted-constrained vs. restricted-free × *similarity to English* × *session*	-1.75	0.89	3.76	0.053
unrestricted vs. restricted-free × *similarity to English* × *session*	-0.98	0.83	1.33	0.249

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4; p values are not adjusted for multiple comparisons. Predictors were coded according to the scheme in Table B1.

Button-Pressing Experiment

Due to convergence issues and to avoid overparameterization, we constrained all correlation parameters to zero. To avoid a singular fit, we removed all second-order random slopes by participant, as well as the random slope by participant for the *restricted-constrained vs. restricted-free* contrast from the model.

Table B4

Button-Pressing Experiment Free-Context-As-Baseline Results

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	-0.56	0.12		
session	-0.22	0.2	1.18	0.276
restricted position	-0.23	0.25	0.84	0.359
restricted-constrained vs. restricted-free	0.35	0.15	5.71	< 0.05 *
unrestricted vs. restricted-free	-0.01	0.22	< 0.01	0.955
restricted position × *session*	0.04	0.41	0.01	0.916
restricted-constrained vs. restricted-free × *session*	-0.69	0.29	5.57	< 0.05 *
restricted-constrained vs. restricted-free × *restricted position*	-0.23	0.29	0.59	0.442
unrestricted vs. restricted-free × *session*	-0.84	0.31	7.39	< 0.01 **
unrestricted vs. restricted-free × *restricted position*	-0.07	0.44	0.02	0.876
restricted-constrained vs. restricted-free × *restricted position* × *session*	1.29	0.58	4.9	< 0.05 *
unrestricted vs. restricted-free × *restricted position* × *session*	0.55	0.61	0.81	0.368

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 3.4.1.1.4; p values are not adjusted for multiple comparisons. Predictors were coded according to the scheme in Table B1.

To probe the nature of the interactions with *session*, we ran follow-up analyses that split the data by day. When analyzing data from each day, we used the same formula as in the main free-context-as-baseline model, but removed *session* and its interactions from the predictors.

Day 1

To avoid overparameterization, we constrained all correlation parameters to zero. To avoid a singular fit, we removed all second-order random slopes plus the random slope of the *restricted-constrained vs. restricted-free* contrast by participant. Results from this model are shown in Table B5.

Table B5

Button-Pressing Experiment Free-Context-As-Baseline Results: Day 1

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	-0.44	0.15	NA	NA
restricted position	-0.25	0.3	0.73	0.393
restricted-constrained vs. restricted-free	0.71	0.18	16.05	< 0.001 ***
unrestricted vs. restricted-free	0.46	0.24	3.60	0.058
restricted-constrained vs. restricted-free × *restricted position*	-0.86	0.36	5.87	< 0.05 *
unrestricted vs. restricted-free × *restricted position*	-0.25	0.48	0.27	0.604

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 3.4.1.1.4; p values are not adjusted for multiple comparisons. Predictors were coded according to the scheme in Table B1.

Day 2

To avoid overparameterization, we constrained all correlation parameters to zero. To avoid a singular fit, we removed all second-order random slopes plus the random slope of the *restricted-constrained vs. restricted-free* contrast by participant. Results from this model are shown in Table B6.

Table B6

Button-Pressing Experiment Free-Context-As-Baseline Results: Day 2

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	*p* value of χ^2 test
(Intercept)	-0.68	0.17		
restricted position	-0.15	0.34	0.21	0.648
restricted-constrained vs. restricted-free	-0.04	0.24	0.02	0.877
unrestricted vs. restricted-free	-0.56	0.32	3.18	0.075
restricted-constrained vs. restricted-free × *restricted position*	0.32	0.47	0.47	0.495
unrestricted vs. restricted-free × *restricted position*	-0.07	0.64	0.01	0.91

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and *p* values for each predictor were obtained via the significance testing procedure described in Section 3.4.1.1.4; *p* values are not adjusted for multiple comparisons. Predictors were coded according to the scheme in Table B1.

Appendix C: Restoration from Sleep

We attempted to explore the connection between slow-wave sleep and questionnaire responses regarding restoration from sleep described in Westerlund et al. (2013). To do this, we analyzed sleep data from the 56 participants who provided full demographic information, recorded at least one night of sleep data using the activity tracker, and responded to the KSQ. In order to temper the influence of outliers on the data, we excluded eight instances in which activity trackers recorded a deep sleep duration that was more than two standard deviations away from the mean—*i.e.*, less than 6 minutes or greater than 139 minutes—leaving 130 sleep instances available for analysis. As in Westerlund et al. (2013), we mapped questionnaire responses onto a 6-point numeric scale[19] and averaged the responses to KSQ questions 9B, 9H, and 9L to form the "restoration from sleep" index (KSQ-RS), We then calculated the Spearman's rank correlation between the KSQ-RS and the number of minutes in deep sleep on each night. We observed no significant relationship between KSQ-RS and deep sleep duration ($\varrho = -0.135$, $p = 0.125$), failing to replicate Westerlund et al. (2013)'s results.

In order to check whether any of the components of the KSQ-RS correlated with deep sleep duration, and to make use of the within-participant variability in sleep data at our disposal, we ran an exploratory mixed-effects linear regression predicting deep sleep duration from 9B, 9H, and 9L, with a random intercept by participant. We used the same software, model convergence procedure, and significance testing criteria as in the speech error model (see described in Section 2.3.2.2 for a description).

[19] Never = 1; Rarely = 2; Sometimes = 3; Often = 4; Mostly = 5; Always = 6

Results from this model are reported in Table C1. The model revealed a significant effect of question 9H, such that participants who reported more instances of not feeling refreshed when waking up had significantly lower sleep durations (β = -6.64, SE = 2.92; $\chi^2(1)$ = 8.97, p < 0.01).

Table C1.

Slow-Wave Sleep Duration Mixed-Effects Model Results

Predictor	Estimate	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	86.80	10.29		
9B—*Difficulties waking up*	-0.14	2.30	3.50	0.061
9H—*Not feeling refreshed when waking up*	-6.64	2.92	8.97	< 0.01 **
9L—*Feeling exhausted when waking up*	2.08	2.91	4.49	0.034

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4.

Appendix D: Analysis of Day 1 Data

In order to determine whether the phonotactic restrictions used by participants affected their productions on Day 1, we ran a version of our model which excluded *session* and all its interactions on data from Day 1 of each experiment.

Experiment 1

To avoid singular model fit, we removed the *restricted-constrained vs. restricted-free* × *similarity to English* random slope from the model. Results are shown in Table D1.

Table D1

Experiment 1 Day 1 Model Results

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	p value of χ^2 test
(Intercept)	0.67	0.13		
restricted-constrained vs. restricted-free	1.19	0.40	7.84	< 0.01 **
unrestricted vs. restricted	-1.53	0.26	23.42	< 0.001 ***
similarity to English	1.39	0.26	20.56	< 0.001 ***
restricted-constrained vs. restricted-free × *similarity to English*	0.10	0.78	0.02	0.874
unrestricted vs. restricted × *similarity to English*	-0.15	0.5	0.09	0.759

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and p values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4; p values are not adjusted for multiple comparisons.

Experiment 2

To avoid convergence issues, we set all correlation parameters to zero. To avoid singular model fit, we removed all second-order random slopes from the model, as well as the random slope by participant for *restricted-constrained vs. restricted-free*.

Table D2

Experiment 2 Day 1 Model Results

Predictor	Estimate (in logits)	Standard Error	$\chi^2(1)$	*p* value of χ^2 test
(Intercept)	0.81	0.15		
restricted-constrained vs. restricted-free	0.74	0.46	2.31	0.128
unrestricted vs. restricted	-0.95	0.28	9.73	< 0.01 **
similarity to training	-0.95	0.3	8.96	< 0.01 **
restricted-constrained vs. restricted-free × *similarity to training*	2.76	0.93	8.53	< 0.01 **
unrestricted vs. restricted × *similarity to training*	-0.63	0.55	1.25	0.264

Note. Only fixed effects are shown in this table. The $\chi^2(1)$ and *p* values for each predictor were obtained via the significance testing procedure described in Section 2.3.2.2.4; *p* values are not adjusted for multiple comparisons.