

Article

Not peer-reviewed version

A Prospective Attempt to Observe a Learning Interaction Between Statistical Learning Experiments

[Peter T. Richtsmeier](#)^{*} and Michelle W. Moore

Posted Date: 2 September 2025

doi: 10.20944/preprints202509.0147.v1

Keywords: statistical learning; frequency; phonology; phonotactics; child speech development



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Prospective Attempt to Observe a Learning Interaction Between Statistical Learning Experiments

Peter T. Richtsmeier ^{1,*} and Michelle W. Moore ^{2,3}

¹ Communication Sciences and Disorders, Oklahoma State University, Stillwater, OK, USA

² Communication Sciences and Disorders, University of Maine, Orono, ME, USA

³ Communication Sciences and Disorders, West Virginia University, Morgantown, WV, USA

* Correspondence: prichtsmeier@yahoo.com

Abstract

This study investigated whether statistical learning from one linguistic input influences learning of a subsequent input in preschool-aged children. The first input comprised word-medial consonant sequences in an "alien" word-learning experiment (e.g., /st/ or /nt/ in alien names). The second input comprised different consonant sequences (e.g., /sk/ and /ns/) in a follow-up "make-believe animal" statistical learning experiment. Thirty-four children were pseudorandomly assigned to one of two word sets in the initial alien experiment and subsequently completed a statistical learning task with manipulated the experimental frequency of the consonant sequences in the make-believe animal names. We hypothesized that prior exposure to /st/ or /nt/ in the alien names would lead to predictable phonological errors—fronting or stopping of /sk/ and /ns/, respectively—in the make-believe animal task. Mixed-effects and ordinal regression analyses revealed no significant effects of alien word set on error patterns or production accuracy for make-believe animal names, though experimental frequency influenced accuracy differently across consonant sequences. These findings suggest that while statistical learning is sensitive to input frequency, cross-experiment interactions may require greater similarity between inputs to manifest. The study contributes to the growing literature on multistream statistical learning and highlights the complexity of generalization in child speech development.

Keywords: statistical learning; frequency; phonology; phonotactics; child speech development

1. Preamble

We report on a phonotactic statistical learning experiment in which children were familiarized with word-medial consonant sequences in nonwords contextualized as make-believe animals. They were then tested on their ability to produce those sequences in a new set of test nonwords. This experiment was preceded by a separate word-learning experiment in which children learned 16 alien names which were also nonwords. The word-learning experiment is reported by Richtsmeier et al. (2024). The two experiments were conducted with the same children because we had hypothesized that the initial alien word-learning experiment would influence the results of the subsequent phonotactic learning experiment. In other words, one hypothesis related to the phonotactic learning experiment was to examine the effects of one experiment on another. This manuscript is specific to that hypothesis.

2. Literature Review

There is a small but growing branch of the statistical learning literature that explores how learners respond to two or more learning inputs, that is, two or more artificial languages. These studies are sometimes referred to as multistream statistical learning studies. In them, both infants and adults who are exposed to two artificial languages with competing statistics tend to learn the

first of the two languages (e.g., Benitez et al., 2020; Bulgarelli & Weiss, 2016; Gebhart et al., 2009; Katz & Moore, 2021; Weiss et al., 2009). For example, Benitez et al. (2020) had infants learning from a continuous stream of syllables, the artificial language, in which only statistics signaled word boundaries (cf. Saffran et al., 1996). The stream began similarly to *pabikutibudogolatupabiku*, where *pa* was always followed by *bi* and then *ku*, such that *pabiku* was a word in this first language. However, the stream switched to a second language with different statistics halfway through, a language in which the syllable *pa* appeared in the middle of the word *mupadi*. Thus, the syllable *pa* only signaled a word onset in the first language. Benitez et al.'s first experiment demonstrated that each language on its own was learnable by infants. However, three subsequent experiments found no evidence of learning when infants were presented with both languages. Benitez and colleagues conclude that statistical learning from two competing or highly overlapping artificial languages is challenging.

There is also evidence that learning from two artificial languages reflects a challenge for young children. Richtsmeier and Goffman (2023) had 4- and 5-year-old children complete two statistical learning tasks. One experiment focused on word-medial consonant sequences such as the /fp/ in /nɪfɹən/, /ʃeɪfɹək/ and /koʊfɹət/. The other experiment focused on prosodic contours with either a trochaic pattern (as in /'mi.fə.pou.bə/) or an iambic pattern (as in /pə.fə.mə.'beɪ/). In both experiments, the targets appeared in either a high frequency or a low frequency familiarization condition. The key finding related to the phonotactic learning experiment: For the children who completed the phonotactic learning experiment first, their accuracy was higher for targets in the high frequency familiarization. In contrast, the children who completed the phonotactic learning experiment second showed higher accuracy for targets in the low frequency familiarization. This latter finding is quite surprising and not consistent with the general expectation for a high-frequency familiarization, specifically, that it should result in greater speech accuracy (Beckman & Edwards, 2000; Edwards et al., 2004). To summarize, Richtsmeier and Goffman studied the same learning targets and the same frequency manipulation in two groups of children, but the experimental frequency effect differed by group and appeared to depend on the presence of a previous statistical learning experiment.

Richtsmeier and Goffman (2023) relate their findings to studies of multistream statistical learning like Benitez et al. (2020), and they interpret their findings as an interaction of, or competition between, two statistical learning tasks. Nevertheless, it remains unclear why children would be affected differently by a second artificial language with properties that are at least superficially unrelated. That is, why would prosodic contours in words with no consonant sequences affect the relationship between speech accuracy and consonant sequence statistics in a second experiment? One possibility follows from the theory that children are hypothesis testers by nature (for example, Gopnik, 2012). The first experiment created expectations, or hypotheses, that children developed regarding what to expect based on familiarization with a set of nonsense words. When the first experiment was a series of prosodic contours, they formed expectations about those contours that they carried into the second experiment. The high experimental frequency input was, in some sense, more in conflict with those expectations because there were more words of that type that did not match the word shapes of the prosodic learning experiment. Thus, experimental frequency did not behave as expected. Nevertheless, it is unclear what caused children to behave differently depending on the order of experiments, partly because the effect was unplanned.

Our goal for this study was to investigate whether children's generalizations from the first of two experiments are predictable enough that their effects on the second experiment could be anticipated. In other words, the goal was to probe for a generalization in children's speech present in the second experiment that could most clearly be attributed to learning in the first experiment.

3. Present Study

In this study, children completed two experiments. The primary question was whether differences in production accuracy in the second experiment would depend on the words they were familiarized with in the first experiment. The first experiment followed the general structure of a

word-learning experiment—children were familiarized with 16 nonwords contextualized as alien names. Beyond learning the aliens’ names, the goal was to familiarize children with a consonant sequence, for example, the /st/ in /mista/ or the /nt/ as in /teintoup/. Two sets of alien names were created, one in which children were familiarized with /st/, the other with /nt/. The sequences /st/ and /nt/ were chosen because they are high frequency in English (Edwards et al., 2004; Vitevitch & Luce, 2004), making learning relatively easy.

In the second experiment, children were familiarized with make-believe animal names containing four word-medial consonant sequences: /sk/, /ns/, /bl/, and /fp/. Two of the consonant sequences, /sk/ and /ns/, were expected to be affected by the first experiment. The sequence /sk/ is similar to the /st/ from the alien names but less frequent in English, or .006 versus .023 per the Phonotactic Probability Calculator (Vitevitch & Luce, 2004). The sequence /ns/ is similar to the /nt/ from the alien names but less frequent in English, or .010 versus .016. Thus, the consonant sequences from the first experiment were generally more frequent, and therefore participants might be more likely to produce /st/ and /nt/ due to ambient language frequency effects. Furthermore, there is a phonological process typical of child speech corresponding to each of the possible errors. Focusing on productions of the make-believe animal names in the second experiment, producing [st] instead of /sk/ would correspond to fronting, and producing [nt] instead of /ns/ would correspond to stopping.

With these error patterns in mind, we made the following hypotheses:

1. Children who complete the alien word-learning experiment and are familiarized with many /st/ exemplars will be more likely to make fronting errors when producing the /sk/ consonant sequence in the subsequent phonotactic learning experiment compared to children familiarized with /nt/ in the alien word-learning experiment.
2. Children who complete the alien word-learning experiment and are familiarized with many /nt/ exemplars will be more likely to make stopping errors when producing the /ns/ consonant sequence in the subsequent phonotactic learning experiment compared to children familiarized with /st/ in the alien word-learning experiment.

4. Methods

The methods reported here were approved by the Institutional Review Board of Oklahoma State University. All participants provided assent, and a parent or guardian provided consent. The data are available through Open Science Framework (OSF): https://osf.io/j8pdf/?view_only=f6b03564b08c4bc9be81fd7bbeeb53e7.

4.1. Participants

Two groups of participants completed the alien word-learning experiment: Participants familiarized with the /st/ alien word set and participants familiarized with the /nt/ alien word set. Children were pseudorandomly assigned to word sets. Of the 46 enrolled in the study, four children met criteria for a speech sound disorder, three did not complete the experiment, two were described by their parent as having a developmental delay, two did not pass the hearing screening, and one was lost due to experimenter error. For the remaining 34 participants, the inclusionary and supplemental developmental measures, as well as the distribution of age and biological sex, are summarized in Table 1. The inclusionary and supplemental measures are reported separately for the two groups along with a statistical comparison. These tests indicate that the two groups were developmentally comparable at the time of participation.

Table 1. Normative Fata for the Participants Separated by the Between-Subjects Factor Word Set.

Task	Word Set 1 - /st/ (<i>n</i> = 19)		Word Set 2 - /nt/ (<i>n</i> = 15)		<i>p</i> (<i>t</i> -test)
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Age in months (range in parentheses)	48.16	(37-59)	48.53	(36-59)	.871
Male-to-female Ratio	9/10		5/10		.409 ^d
Goldman-Fristoe Test of Articulation-2 ^a	108.63	(9.93)	106.12	(8.59)	.425
Narrative Language Screener (out of 28)	6.68	(3.43)	8.12	(4.18)	.655
Test of Early Grammatical Impairment Screener (percentage) ^b	67.41	(25.09)	77.26	(20.90)	.217
Preschool Test of Nonverbal Intelligence ^{a, b}	109.83	(18.71)	109.69	(17.63)	.981
Name-writing ability (out of 4)	1.95	(1.13)	2.12	(0.99)	.636
Flanker overall accuracy (percentage)	70.51	(16.59)	65.74	(15.55)	.381
Nonword repetition overall accuracy (percentage) ^c	73.40	(13.55)	74.16	(16.05)	.839
Parental rating of speech (out of 10)	6.24	(1.68)	6.80	(2.04)	.381
Gross motor functioning (out of 10)	6.92	(1.55)	6.17	(1.36)	.152
Fine motor functioning (out of 10)	6.66	(1.46)	6.27	(1.24)	.422
Socioeconomic status relative to neighbors (out of 10)	6.22	(1.39)	5.63	(1.31)	.223

^a The reported means are of standard scores with an estimated population mean of 100 and standard deviation of 15. ^b Due to irregularities in administration of these tests, the results were not used for inclusion in or exclusion from the study. ^c Due to experimenter error, some children did not complete the nonword repetition task. The correct sample sizes for that task are as follows: Word Set 1 - /st/, *n* = 18; Word Set 2 - /nt/, *n* = 15. ^d The reported *p*-value comes from a chi-squared test, $\chi^2(1) = 0.68$.

Out of the 34 children analyzed, 32 were monolingual English speakers. One child's family reported Hindi as a second language, and another child's family reported Vietnamese as a first language, but both families considered English to be the child's primary language during the experiment. All 34 children met the criteria for typical speech and hearing, with articulation scores of 85 or higher on the Goldman Fristoe Test of Articulation, Second Edition (GFTA-2, Goldman & Fristoe, 2000, *M* = 107.3, Range = 88-126), and they all passed a hearing screening in both ears at 25 dB for pure tones of 1000, 2000, and 4000 Hz. Data from two additional standardized tests were collected but not used for inclusion criteria: the Test of Early Grammatical Impairment (TEGI, Rice & Wexler, 2001, *M* = 71.6%, Range = 20%-100%) and the Primary Test of Nonverbal Intelligence (PTONI, Ehrler & McGhee, 2008, *M* = 109.3, Range = 70-149). These tests showed inconsistencies, with about 50% of TEGI items unscorable and more than half of PTONI scores outside the 1 SD range (4 children scored below 85, and 17 scored at or above 115). These inconsistencies likely reflect challenges in achieving consistent and reliable administration of these tasks, so no child was excluded based on these results.

Additional measures were collected to assess other areas of cognitive and linguistic development. The CUBED Narrative Language Screener (Petersen & Spencer, 2016) assessed children's ability to retell brief narratives. Children heard and retold three narratives from the Preschool Fall set, scored for linguistic and narrative content out of 28 points. Only raw scores are reported, as many participants were likely too young for preschool, and their school status was unknown. Name-writing ability, an early literacy skill indicator, was assessed using a four-point scale (Puranik et al., 2014). Participants also completed a flanker task to assess response inhibition, a component of executive functioning (McDermott et al., 2007), and a nonword repetition task to assess short-term and long-term phonological memory (Moore, 2018). Parents rated their child's speech accuracy, gross motor development, fine motor development, and the family's socioeconomic status relative to their neighbors. These supplementary measures were not included in any analyses. The inclusionary and supplementary developmental measures, along with distributions of age and biological sex, appear in Table 1.

4.2. Materials

The materials from the alien word-learning experiment are reported by Richtsmeier (2024) and are repeated here in Table 2. Each word set contained four base forms, three of which were single syllables and one of which was disyllabic. Each base word was combined with each of three suffixes— /-təf/, /-toop/, and /-ta/—to create 12 additional words, or 16 total per word set. In the suffixed words, the suffix carried stress. The single-syllable base words ended with a coda consonant so that, when combined with the suffixes, the consonant sequence /st/ was made for Word Set 1 and the sequence /nt/ was made for Word Set 2. These consonant sequences represented the generalizations that participants were expected to carry over to the phonotactic learning experiment. One disyllabic word was included in each word set to match the stress pattern on the phonotactic learning words, but when the suffix was added, no consonant sequence was created.

Table 2. The Two Sets of Nonsense Words Developed for the Alien Word-Learning Experiment.

	Base wordform	Base + -təf	Base + -toop	Base + -ta
Word Set 1 - /st/	tas	tastef	tastoop	tasta
	mis	mistef	mistoop	mista
	deis	deistef	deistoop	deista
	zuder	zudertef	zudertoop	zuderita
Word Set 2 - /nt/	tem	teintef	teintoop	teinta
	man	mantef	mantoop	manta
	poun	pountef	pountoop	pounta
	tuker	tukertef	tukertoop	tukerta

The 12 words from the phonotactic learning experiment (hereafter “phonotactics words”) appear in Table 3 below, including three familiarization words and one test word per target consonant sequence. For each target, three familiarization items and one test item were created. The familiarization words in italics were used for the low experimental frequency condition. Each word was associated with a colorful make-believe animal (Ohala, 1999). The phonotactic words featured unique initial and final syllables, with stress placed on the first and last syllables, while the middle syllable remained unstressed. Each word was produced by four different speakers, two males and two females. Participants were recorded in a sound booth where they were instructed to articulate each word clearly. The speakers used a Southern American English dialect, which typically merges the vowels /ɪ/ and /ɛ/ before nasals; however, this phonetic environment was not present in any of the words. The recordings were later edited to remove breaths and tongue clicks, extracted as separate sound files with 50 ms of silence before and after each item, and their intensity was adjusted to 70 dB using Praat software (Boersma & Weenink, 2021). As described above, the phonotactics words containing /sk/ were intended to be similar to Word Set 1 with /st/ from the alien word-learning experiment, allowing for fronting errors. The phonotactics words containing /ns/ were intended to be similar to Word Set 2 with /nt/, allowing for stopping errors.

Two targets were assigned to a low experimental frequency condition and two to a high experimental frequency condition. In the low experimental frequency condition, children heard a target in just one item spoken by the four talkers. In the high experimental frequency condition, children heard all three items spoken by the four talkers. Experimental frequency was therefore a within-subjects manipulation and primarily a difference of word-type frequency (Richtsmeier, 2016). To counterbalance the assignment of items to experimental frequencies and items to make-believe animals, four lists were created. Each list had a different set of high and low experimental frequency targets.

Table 3. The Learning Targets (Word-Medial Consonant Sequences), Familiarization Words, and Test Words for the Phonotactic Learning Experiment.

Target Consonant Sequence	Familiarization Words			Test Words
/ns/	sæn.se.vaʊn	noʊn.si.kik	fun.sə.tʃɪm	tin.so.paɪz
/sk/	bɪs.kə.liθ	vʊs.kə.reb	mʊs.ke.dʒɪf	dʌs.ku.nʌʃ
/fp/	wɛf.po.laɪg	zʌf.pə.tʃas	dʒʊf.pə.taʊθ	bɛf.pu.jʌg
/bl/	mɪb.lu.tʃæv	deɪb.lə.ruʊŋk	kɪb.le.jʊp	sɔɪb.li.haʊf

4.3. Procedure

Throughout the study, for both the alien word-learning experiment and the phonotactic learning experiment, children were seated in front of a computer and speakers. The alien word-learning experiment was always completed first, across two sessions. Participants completed four familiarization conditions during the first session and a post-test during the second session. During the four familiarization blocks in the first session, children were exposed to all 16 alien names two times each. Two familiarization conditions emphasized perceptual learning where children passively listened to the alien names. Two familiarization conditions emphasized production learning where children repeated the alien names. This manipulation of a learning modality factor was crossed with a semantic depth factor. For semantic depth, two familiarization conditions involved exposure only to the alien names and pictures (Gupta et al., 2004). Two other familiarization conditions involved additional semantic cues for the suffixed aliens. Following each familiarization block, participants completed a test block that included two referent identification tasks and a naming task. In the post-test, participants completed the test block for a fifth time but without a familiarization. Additional details can be found in Richtsmeier et al. (2024). Regarding exposure to the target consonant sequences, each list contained 9 alien names with the target sequence. Over the course of the four familiarizations, participants heard their target sequence 72 times and had 36 opportunities to produce those sequences.

The phonotactic learning experiment was completed during the second session immediately after the alien word-learning post-test. This was intended to maximize the chances that generalizations about consonant sequences made during the alien word-learning task might influence learning in the phonotactic learning experiment.

The experimental design was similar to infant statistical learning studies, beginning with a passive familiarization phase. Speech production accuracy was measured during a test immediately following the familiarization. At the start and just before the familiarization, children were instructed that they would hear the names of make-believe animals and their task was to watch and listen. The familiarization phase included 8 low-frequency trials (2 targets × 1 word × 4 talkers) and 24 high-frequency trials (2 targets × 3 words × 4 talkers). The 32 trials played automatically in a predetermined, pseudorandom order, ensuring no item was repeated consecutively. Each of the four lists had a different pseudorandom order. The familiarization was divided into two blocks of 16 trials, with a short break in between to reinforce the instructions to watch and listen. Each trial lasted about 1.8 seconds, and the entire familiarization phase took about 1 minute.

After the familiarization, children were told they would see new make-believe animals, hear their names, and their task was to repeat these new names. Thus, the productions made during the test can be considered imitations following a model. Children had 5 opportunities to produce each test item, totaling 20 opportunities. Most children completed the majority of the 20 productions (M = 18.0, SD = 3.02, range 7–20).

4.4. Analyses

4.4.1. Primary Analysis

The main hypotheses were created with respect to the effects of the two word sets from the alien word learning study on errors made during the phonotactic learning experiment. Briefly, children who were exposed to alien names like /tʌstoup/ and /mista/ were predicted to make fronting errors when producing the test item /daskunʌʃ/ (/daskunʌʃ/ → [dastunʌʃ]). Children who were exposed to alien names like /teintoup/ and /manta/ were predicted to make stopping errors when producing the test item /tinsopaɪz/ (/tinsopaɪz/ → [tintopaɪz]).

The dependent measure was the presence of a fronting or stopping error. Errors were coded specific to the /k/ in the /sk/ consonant sequence in /daskunʌʃ/, and the /s/ in the /ns/ consonant sequence in /tinsopaɪz/. Fronting errors on the /k/ in /sk/ and stopping errors on the /s/ in /ns/ were coded as hits. All other errors were coded as misses, resulting in a binomial dependent variable. During data coding, it was noted that some errors could be considered “fronting” or “stopping” other than [t]. For example, a production of /daskunʌʃ/ as [dasdunʌʃ] could reflect fronting in combination with word- or syllable-initial voicing. Therefore, transcriptions of the target /k/ in /daskunʌʃ/ that had the error sounds [t], [d], or [n] were coded as fronting. Transcriptions of the target /s/ in /tinsopaɪz/ that had the error sounds [t] or [d] were coded as stopping. Children’s productions were transcribed by the first author, and a research assistant separately transcribed productions from eight participants. The research assistant was not informed about the purpose of the study. Phoneme-to-phoneme reliability was 80.34% and was considered acceptable.

Errors were coded with the type (fronting or stopping) being one fixed effect and word set as a second fixed effect. Data were entered into a mixed effects logit model using the *lme4* package in R (Bates et al., 2015) for a binomial distribution, with random intercepts for participants.

4.4.2. Secondary Analysis

In addition to the analysis of fronting and stopping errors, a potential interaction with the word set from the alien word-learning experiment was also considered for the effects of experimental frequency. This analysis follows from the results reported by Richtsmeier and Goffman (2023) in which the interaction of two experiments was observed in the behavior of high experimental frequency versus low experimental frequency. The data were transcriptions of the four target sequences (/sk/, /nt/, /bl/, and /fp/) scored following conventions from Edwards et al. (2004). For each consonant in a sequence, possible scores were 0 (missing consonant), 1 (2+ features off from target consonant), 2 (1 feature off from target consonant), or 3 (correct production). Thus, the possible range of scores for each sequence was 0-6. The data were entered into an ordinal regression using the *ordinal* package in R (Christensen, 2023). Initial model testing began with a baseline model that contained experimental frequency and word set as fixed effects and participant as a random effect. The alternative model allowed for an interaction of experimental frequency and word set, and the baseline and alternative models were compared using the *anova* function in R for model comparison.

5. Results

With respect to the /k/ in the /sk/ consonant sequence in /daskunʌʃ/, and the /s/ in the /ns/ consonant sequence in /tinsopaɪz/, a total of 93 errors were recorded across all 34 participants and were included in the analyses. There were 53 errors coded for the /k/ in /daskunʌʃ/. Of those, 32 errors were classified as fronting, and the error was transcribed as [t] 30 times and as [d] 2 times. There were 40 errors coded for the /s/ in /tinsopaɪz/. Of those, 12 errors were classified as stopping, and the error was transcribed as [t] all 12 times. For the statistical analysis of errors, there were no significant effects (all *ps* > .400). See Table 4 for a model summary.

Table 4. Mixed Effects Model Results for the Analysis of Fronting and Stopping Errors.

Fixed Effects				
	Estimate	Standard Error	z-value	p
Intercept	0.94	1.18	0.79	.430
Word Set	-0.32	0.75	-0.42	.672
Error Type	-0.93	1.62	-0.57	.567
Word Set*Error Type	-0.35	1.05	-0.34	.738
Random Effects				
			Variance	Standard Deviation
Participant (Intercept)			0.47	0.69
Model Fit				
AIC	BIC	Log Likelihood	Deviance	df residual
127.9	140.6	-59.0	117.9	88
N total observations = 93				
Model equation: Error ~ 1 + Word Set * Error Type + (1 Participant)				

For the secondary analysis, transcriptions were entered into a baseline ordinal regression in R software using the *ordinal* package, with experimental frequency and word set as fixed effects and participant as a random effect. That model was compared with an alternative model with an interaction term for experimental frequency and word set (all *ps* > .200). Neither model included a significant main effect, and a comparison of the two models’ likelihood ratios using the *anova* function was not significant (LR = 0.70, *df* = 1, *p* = .402). A corresponding graph can be seen in Figure 1 below.

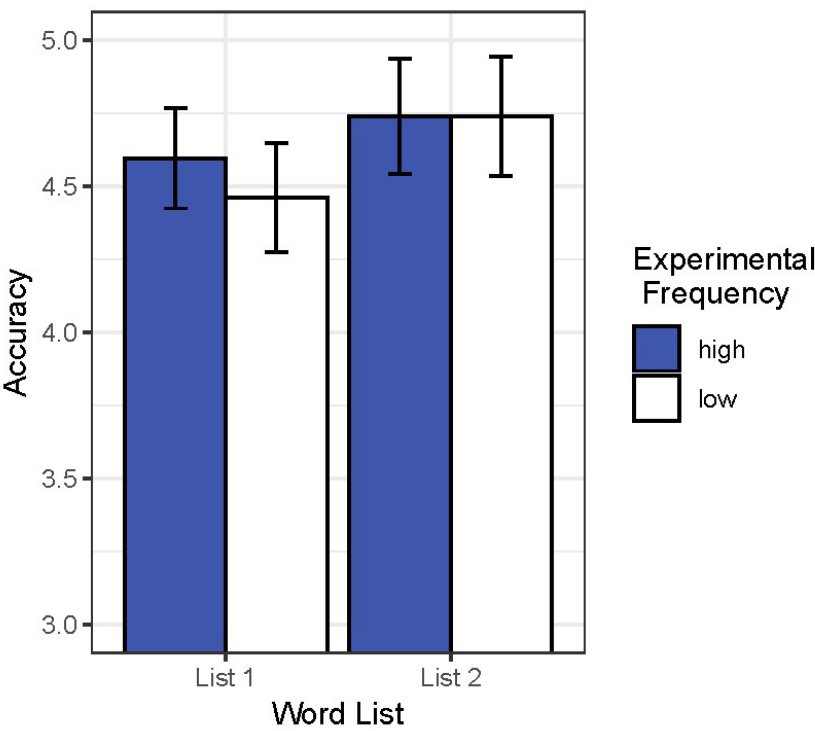


Figure 1. Average Accuracy for the Target Word-Medial Consonant Sequences out of a Possible 6 Points.

In a follow-up, consonant sequences were included as a factor based on the hypothesis that they might interact with the word sets. Specifically, Word Set 1 was expected to affect /sk/, and Word Set 2 was expected to affect /ns/. Thus, ordinal regressions were developed. The baseline model contained word set, experimental frequency, and consonant sequence as fixed effects and participant as a random effect. The alternative model included an interaction of experimental frequency and

consonant sequence. Here, the alternative model provided a significantly better fit for the data ($LR = 18.13$, $df = 3$, $p < .001$) and appears in Table 5. There was a significant main effect of experimental frequency ($p = .001$), indicating that participants were generally more accurate when producing high experimental frequency sequences. However, accuracy varied by consonant sequence, and compared to the effect of experimental frequency on /bl/, it behaved differently for /ns/, $p = .002$, and /sk/, $p < .001$. There was no effect of word set, however, $p = .700$. To summarize, when the target four consonant sequences were considered as separate levels of a fixed effect, the effect of experimental frequency depended on the specific target.

Table 5. Summary of the Alternative Ordinal Regression Model with Word Set, Experimental Frequency, and Consonant Sequence as Fixed Effects. Consonants Sequences were Compared to /bl/.

Fixed Effects	Estimate	Standard Error	z-value	p-value
Word set	0.20	0.53	0.39	.700
Experimental Frequency	-1.22	0.39	-3.18	.001*
/fp/	1.33	0.40	3.35	<.001*
/ns/	1.03	0.41	2.49	.013*
/sk/	0.33	0.34	0.96	.337
Experimental Frequency × /fp/	0.32	0.56	0.57	.571
Experimental Frequency × /ns/	2.08	0.68	3.07	.002*
Experimental Frequency × /sk/	2.29	.58	3.95	<.001*
Model Summary				
Observations	612			
Log Likelihood Ratio	-743.64			
AIC	1517.27			

* $p < .05$.

To better understand the interaction of experimental frequency and consonant sequence, separate ordinal regressions for each consonant sequence were developed with word set, experimental frequency, and production number as fixed effects and participant as a random effect. Baseline models included only main effects. Alternative models included an interaction between word set and experimental frequency, but ANOVAs comparing the initial and alternative models were not significant (all p -values $> .170$), so only the baseline models are interpreted. Summaries of those results appear in Table 6 below. In sum, the varying effects of experimental frequency across consonant sequences did not depend on word set.

Table 6. Summary of Ordinal Regressions for Each Consonant Sequence. Only the Main Effects of Word Set, Experimental Frequency, and Production Number were Included in these Analyses.

Consonant Sequence	Word set		Experimental Frequency		Production Number		No. Obs.	Log Likelihood	AIC
	z-score	p	z-score	p	z-score	p			
/bl/	0.28	.777	-3.75	<.001*	1.88	.060	157	-189.40	398.80
/fp/	0.51	.614	-0.26	.797	1.00	.320	148	-175.85	371.69
/ns/	-0.27	.790	2.42	.016*	-0.27	.786	150	-143.71	307.41
/sk/	0.17	.868	0.95	.343	-0.81	.420	157	-175.00	365.99

6. Discussion

The purpose of this manuscript was to describe the hypotheses, design, analyses, and findings specific to a possible interaction of two experiments: a word-learning experiment (Richtsmeier et al., 2024) and a phonotactic statistical learning experiment. Children first completed the alien word-learning experiment in which they were either familiarized with many alien names containing /st/, or with many alien names containing /nt/. Next, children completed the phonotactic learning experiment in which they were familiarized with and produced four consonant sequences, including /sk/ and /ns/, appearing in make-believe animal names. Based on previous findings from multistream statistical learning, as well as a reported interaction of two experiments completed by similarly aged

children, exposure to the /st/ and /nt/ consonant sequences in the alien names was hypothesized to lead to fronting and stopping errors, respectively, on the /sk/ and /ns/ consonant sequences in the make-believe animal names.

Although children did make both fronting and stopping errors in the phonotactic statistical learning experiment, the errors did not reveal an effect of the familiarization from the alien word-learning experiment. Additionally, the familiarization from the alien word-learning experiment did not appear to affect how experimental frequency influenced production accuracy in the phonotactic learning experiment. In sum, there were no observed effects of one experiment on the other.

There is substantial evidence that two statistical learning inputs can interact. This has been shown most clearly by Gebhart et al. (2009), Weiss et al. (2009), and other multistream statistical learning studies in adults. In these experiments, the two inputs, or artificial languages, are composed of the same syllables, which appear sequentially and often with no cue to distinguish them, and they are therefore virtually identical other than the transitional probabilities between syllables. Thus, when two inputs are highly similar, there is strong evidence that humans struggle to keep them separate.

There is also evidence of an interaction between less similar inputs, as was shown by Richtsmeier and Goffman (2023), whose participants completed sequential statistical learning experiments with prosody- and phonotactics-oriented inputs. Participant behavior in that phonotactic learning experiment depended on the order of the two experiments, although the results from the prosody learning experiment did not (for analyses of the prosodic learning experiment, see Richtsmeier & Goffman, 2021). Therefore, it appears that the more that two inputs differ, the more reason to believe that learning from one input will have little or no effect on learning from the other. One possible explanation for the null results reported here is that the two experiments were sufficiently different—in terms of item phonology, item morphology, visual referents, etc.—so it was unlikely that the alien word-learning experiment would influence the results of the phonotactic learning experiment. Of course, the experiments were designed to be similar in terms of exposing participants to word-medial consonant sequences. For now, we consider it an open question as to which factors, and what degree of similarity, are required for such an interaction to occur. Generally, more similarity should increase the chances of an interaction.

Setting aside the issue of degree of similarity, this study is valuable as one of a relatively small number of studies examining the nature of an interaction between experiments. Despite the clear evidence of an interaction in virtually all multistream statistical learning studies, it is unclear how those interactions should be interpreted. The primacy effect, or the tendency for participants to respond consistently with learning from the first of two inputs, could be happening for several reasons. Here, we consider two. First, participants may fail to store the second input under most circumstances, that is, the breakdown occurs in memory via forgetting or interference. Second, participants may learn patterns from the first input and then apply them to the test for later inputs. This second explanation relies more on learning theory, but it has not been considered to date.

We consider it unlikely that participants forgot the first input. This notion is supported by findings such as those of Katz and Moore (2021), who found an apparent interaction of two statistical learning inputs in adults even though their participants were exposed to the second input at least 30 days after the first (note that there was no carryover effect, or interaction of the two artificial languages, observed in their sample of third- through fifth-grade children, though this was unsurprising because the overall learning effect was marginal and highly variable in this sample of children). Thus, memories from a statistical learning task appear unlikely to be forgotten quickly. Here, we attempted to test the second possibility, that learners attempt to apply the first input when learning the second input. Participants were taught a specific phonotactic pattern in the alien word-learning experiment, and then we probed for learning of that pattern in the phonotactic learning experiment. Given the null results, the results are not informative about either of the explanations above for results in multistream statistical learning experiments.

Despite a lack of support for the hypothesis that two experiments interact, we believe that there is value in conducting further studies with two competing or overlapping statistical inputs. Consider the two error patterns that were probed here: fronting and stopping. Both patterns are commonly observed in early speech development, yet children do not hear adults make these types of errors. Stemberger and Bernhardt (1999) argue that these errors could reflect natural language frequency, specifically, the frequency of the error sound. That is, /t/ is more frequent than both /k/ and /s/ and is therefore naturally the result of the fronting and stopping phonological processes. However, why would children ever conflate two sounds that are contrastive in a language? The statistical learning paradigm offers a way to explore the explanation that the errors may arise from a mistaken generalization. If children hear /t/ many times in succession or hear /t/ in a pattern such as at the start of the second syllable of three words in a row, they may erroneously generalize use of /t/ in productions where it would not normally occur. Given the frequency of /t/, children likely have many opportunities to form such an erroneous generalization. This explanation does not require frequency to be the only factor responsible for phonological processes like fronting and stopping (for example, Jarosz, 2017), but it does naturally lead to tests of such explanations using the statistical learning paradigm. Furthermore, so long as future studies establish an interaction between experiments, the paradigm may also be useful for testing other explanations for phonological processes, including underspecification (Stemberger, 1991), or examining rare phonological processes such as backing (/t/ → [k]). Thus, continued exploration of experimental interactions in the study of child speech may be fruitful.

Disclosure of AI Tools: The authors report that AI was used for writing the Materials and Procedures sections of the Methods to vary the wording relative to another manuscript describing the same data but different hypotheses. No data were created through AI, and all statistical analyses were conducted by the first author.

Data Availability Statement: The data analyzed in this manuscript are available through Open Science Framework (OSF) via this link: https://osf.io/j8pdf/?view_only=f6b03564b08c4bc9be81fd7bbeeb53e7.

Acknowledgments: The authors extend gratitude to Megan McLeod, Bridget DeLeon, Lauren Lamer, and Anna Edmondson for their assistance with participant recruitment, data collection, data processing, and reporting to the participants' families. Additional thanks to Sarah Slagle for reliability scoring. This research was not supported by grant funds from agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest: The authors declare no conflicts of interest other than the desire to publish for career advancement. This research was not supported by external grant funding.

References

1. Beckman, M. E., & Edwards, J. (2000). Lexical frequency effects on young children's imitative productions. In M. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 207–217). Cambridge University Press.
2. Benitez, V. L., Bulgarelli, F., Byers-Heinlein, K., Saffran, J. R., & Weiss, D. J. (2020). Statistical learning of multiple speech streams: A challenge for monolingual infants. *Developmental Science*, 23(2), e12896. <https://doi.org/10.1111/desc.12896>
3. Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer [Computer software].
4. Bulgarelli, F., & Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1621. <https://doi.org/10.1037/xlm0000263>
5. Christensen, R. H. B. (2023). ordinal—Regression Models for Ordinal Data (Version R package version 2023.12-4.1) [Computer software]. <https://CRAN.R-project.org/package=ordinal>
6. Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2), 421–436. [https://doi.org/10.1044/1092-4388\(2004/034\)](https://doi.org/10.1044/1092-4388(2004/034))

7. Ehrler, D. J., & McGhee, R. L. (2008). PTONI: Primary Test of Nonverbal Intelligence [assessment instrument]. Pro-Ed.
8. Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, 33(6), 1087–1116. <https://doi.org/10.1111/j.1551-6709.2009.01041.x>
9. Goldman, R., & Fristoe, M. (2000). Goldman-Fristoe Test of Articulation-2 [assessment instrument]. Pearson.
10. Gopnik, A. (2012). Scientific Thinking in Young Children: Theoretical Advances, Empirical Research, and Policy Implications. *Science*, 337(6102), 1623–1627. <https://doi.org/10.1126/science.1223416>
11. Gupta, P., Lipinski, J., Abbs, B., Lin, P.-H., Aktunc, E., Ludden, D., Martin, N., & Newman, R. (2004). Space aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behavior Research Methods, Instruments, & Computers*, 36(4), 599–603. <https://doi.org/10.3758/BF03206540>
12. Jarosz, G. (2017). Defying the stimulus: Acquisition of complex onsets in Polish. *Phonology*, 34(2), 269–298. <https://doi.org/10.1017/S0952675717000148>
13. Katz, J., & Moore, M. W. (2021). Phonetic Effects in Child and Adult Word Segmentation. *Journal of Speech, Language, and Hearing Research*, 64(3), 854–869. https://doi.org/10.1044/2020_JSLHR-19-00275
14. McDermott, J. M., Pérez-Edgar, K., & Fox, N. A. (2007). Variations of the flanker paradigm: Assessing selective attention in young children. *Behavior Research Methods*, 39(1), 62–70. <https://doi.org/10.3758/BF03192844>
15. Moore, M. W. (2018). Consonant age of acquisition effects are robust in children's nonword repetition performance. *Applied Psycholinguistics*, 39(5), 933–959. <https://doi.org/10.1017/S014271641800005X>
16. Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*, 32(6), 397–422. [https://doi.org/10.1016/S0021-9924\(99\)00018-0](https://doi.org/10.1016/S0021-9924(99)00018-0)
17. Petersen, D. B., & Spencer, T. D. (2016). CUBED. Laramie, WY: Language Dynamics Group. <https://languagedynamicsgroup.com/cubed-overview/>
18. Puranik, C. S., Schreiber, S., Estabrook, E., & O'Donnell, E. (2014). Comparison of name-writing rubrics: Is there a gold standard? *Assessment for Effective Intervention*, 40(1), 16–23. <https://doi.org/10.1177%2F1534508413502390>
19. Rice, M., & Wexler, K. (2001). Rice Wexler Test of Early Grammatical Impairment [assessment instrument]. Hove.
20. Richtsmeier, P. T. (2016). Phonological and semantic cues to learning from word-types. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1). <https://doi.org/10.5334/labphon.39>
21. Richtsmeier, P. T., Gladfelter, A., & Moore, M. W. (2024). Contributions of speaking, listening, and semantic depth to word learning in typical 3- and 4-year-olds. *Language, Speech, and Hearing Services in Schools*, 1–14. https://doi.org/10.1044/2024_LSHSS-23-00105
22. Richtsmeier, P. T., & Goffman, L. (2021). A learning interaction between statistical learning experiments. *BEHAVIORAL SCIENCES*. <https://doi.org/10.20944/preprints202105.0777.v1>
23. Richtsmeier, P. T., & Goffman, L. (2023). Statistical learning of phonotactics by children can be affected by another statistical learning task. *Applied Psycholinguistics*, 44(6), 1124–1142. <https://doi.org/10.1017/S0142716423000449>
24. Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
25. Stemberger, J. P. (1991). Radical underspecification in language production. *Phonology*, 8(1), 73–112. <https://doi.org/10.1017/S0952675700001287>
26. Stemberger, J. P., & Bernhardt, B. H. (1999). The emergence of faithfulness. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 417–446). Erlbaum.
27. Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3), 481–487. <https://doi.org/10.3758/BF03195594>
28. Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development*, 5(1), 30–49. <https://doi.org/10.1080/15475440802340101>

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.