 doi:10.20944/preprints202005.0230.v1

Order in the Statistical Learning of Phonotactics

Peter T. Richtsmeier, Ph.D., CCC-SLP

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

Physical Address: 042 Murray Hall, Oklahoma State University, Stillwater, OK, 74078

Email Address: prichtsmeier@yahoo.com

Dr. Richtsmeier is the corresponding author.

 doi:10.20944/preprints202005.0230.v1

Abstract

A premise of statistical learning research is that learners attend to and learn the frequencies of repeating or co-occurring elements in the input. When the input is a series of words, participants readily learn the frequencies of phoneme sequences, that is, to learn phonotactic frequencies. Inherent to the concepts of both frequency and phonotactics is order, or the temporal structure of the input. Order is similarly inherent to statistical learning, yet the effect of order on statistical learning is not well understood. In the present study, adult participants learned the relative frequencies of eight item-medial consonant sequences, for example, the /mk/ in /nʌmkət/. Across five ordering conditions, both familiarization and test stimuli were independently ordered and randomized, thus allowing for a relatively broad search for order effects in an established statistical learning paradigm. Participants learned the target frequencies equivalently across the five ordering conditions, indicating no modulating effect of order. Nevertheless, participants also approached the task by applying idiosyncratic, structured orders to their responses. The result is an unexpected but robust effect of order. Both the results and the design of the study also allow for increased integration of statistical learning with memory and other aspects of cognition. (198 words)

Keywords: statistical learning; phonotactics; order effects; working memory; mixed effects modeling

 doi:10.20944/preprints202005.0230.v1

Order in the Statistical Learning of Phonotactics

Introduction

Phonotactics are restrictions on sounds, especially consonants, in terms of where and in what order they can appear in a syllable or word. Many phonotactic restrictions are relative and probabilistic, reflecting the statistics of a language. For example, native speakers of English are sensitive to the frequencies of word-initial onset clusters. They know the /kl/ in 'clean' is a relatively frequent onset cluster of English, the /gl/ in 'glee' is somewhat less frequent, and /ʃl/ is rare, appearing only in borrowings like *shlep*, /ʃlɛp/ 'carry a heavy load', from Yiddish (see for example, Albright, 2009; Bailey & Hahn, 2001; Pierrehumbert, 1994). Probabilistic phonotactics are also possible across syllables. In Māori, for example, only CV syllables are allowed, but consonants vary in frequency across the onsets of multisyllabic words. Stops are frequently the first onset of a word, the rhotic /r/ is frequently the second onset, and sequences of homorganic onsets are rare (Rácz, Hay, Needle, King, & Pierrehumbert, 2016). Finally, probabilistic phonotactics are highly relevant to words, as when low probability sequences help learners identify word boundaries. The consonant sequence /fb/ does not occur within English words, and therefore indicates a likely word boundary. English speakers use that information to parse /lifbloʊwə˞/ as 'leaf blower' and not 'lee fblower'. Weber and Cutler (2006) showed that low probability sequences like /fb/ allow L1 and L2 speakers of English to readily find real English words like *leaf* in larger phonetic sequences like /lifbɛp/. Higher probability sequences that are common in English words, like /fl/ in the sequence /liflɛp/, make spotting *leaf* more difficult. The present study concerns the learning of probabilistic phonotactic sequences, and how learning these sequences may be affected by the order of the input. Both learning and order are discussed in detail below.

 doi:10.20944/preprints202005.0230.v1

Phonotactics and Learning

There is a relatively straightforward link between phonotactic probabilities and learning. If English speakers know that /fb/ typically signals a word boundary and /fl/ an onset cluster, they must have learned these patterns via exposure to the English language. Evidence of such learning is widespread, including indirect evidence from statistical learning tasks. In one such task, Adriaans and Kager (2017) familiarized Dutch speakers with consonant frames (for example, p_d_g_) in a continuous, synthesized speech stream. Vowels were added to the frames randomly, meaning that there were never repeated "words". At test, participants distinguished between novel items that matched or did not match the familiarization frame, indicating that they had learned the phonotactic patterns, even without the assistance of word boundaries or repeated words.

Similar evidence for probabilistic phonotactic learning comes from Richtsmeier (2011), who found that adult participants develop preferences for consonant sequences that are heard frequently during a familiarization. Participants were familiarized with CVCCVC nonwords that were contextualized as Martian animal names and that contained one of eight target medial consonant sequences. For example, the familiarization nonwords /nʌmkət/, /gumkən/ and /ʃamkən/ all contained word-medial /mk/. Some target sequences appeared in three nonwords during the familiarization. Others appeared in just a single nonword, a manipulation of the experimental frequency of the eight target sequences. Following familiarization, participants heard test nonwords containing the same eight sequences, for example, /mk/ in /dɪmkəs/, /kəɪmkət/, and /sæmkəs/. Participants rated these test nonwords using a 7-point Likert wordlikeness scale ranging from "definitely not a Martian animal name" to "a great Martian animal name". The results revealed a high experimental frequency advantage: Test nonwords

 doi:10.20944/preprints202005.0230.v1

were rated as more Martian-like if they contained sequences that appeared in three familiarization nonwords. Thus, the experimental frequency manipulation resulted in phonotactic probabilities within the experiment. When sequences like /mk/ were frequent in the familiarization, test nonwords like /dɪmkəs/ were rated higher than when /mk/ was infrequent (see also Denby, Schecter, Arn, Dimov, & Goldrick, 2018; Onishi, Chambers, & Fisher, 2002).

To succeed in the tasks reported by Adriaans and Kager (2017) and Richtsmeier (2011), learners must attend to the frequencies of repeated elements appearing in the experimental input. However, there is increasing evidence that statistical learning involves sensitivity to more than the raw frequencies of the input. Beyond frequency counts, emerging research suggests that input order, or the order in which familiarization items occur over time, is an important component of statistical learning. Below, several effects of order reported in the literature are broken down into two broad categories: cases where order inhibited learning and cases where it facilitated learning. Given that order results in divergent influences on learning, the present study is motivated by the assumption that order is worth studying in greater detail.

Order Sometimes Inhibits Learning

Several lines of existing research suggest that order can inhibit learning. For example, it may create local or spurious generalizations. In a study with infants, Gerken and Quam (2017) found that orders with unintended local generalizations could outweigh the more general statistics of the familiarization. Eleven-month-olds were exposed to one of two artificial languages in which the two consonants of novel CVCV words were related, either by place of articulation (*poba* contains two labials) or voicing (*dova* contains two voiced consonants). All infants were familiarized with 24 words, but some infants heard the words in an order that

 doi:10.20944/preprints202005.0230.v1

allowed for local generalizations, for example, when two or more adjacent words started with the same consonant. Thus, the local generalizations did not contradict the place or voicing generalizations, but they allowed for learning of something more narrow like, "/p/ is a common onset in this input." When local generalizations were present, infants did not appear to learn the more global phonological rules for place of articulation or voicing. When those local generalizations were removed, however, infants did learn the place and voicing rules. Gerken and Quam conclude that order can create local or unintended generalizations in an input that prevents learning of more diffuse patterns.

Certain types of order have also consistently inhibited learning in studies of statistical word segmentation. In statistical word segmentation tasks, participants listen to a continuous stream of syllables like *bidakupadotigolabubidakugolabu*. Some syllables always occur in a sequence, such as *bi*, *da*, and then *ku*, meaning that *bidaku* functions like a word in the stream (Saffran, Aslin, & Newport, 1996). In multistream statistical segmentation, participants are exposed to two different syllable streams, each stream containing its own transitional probabilities (Bulgarelli, Benitez, Saffran, Byers-Heinlein, & Weiss, 2017; Bulgarelli & Weiss, 2016; Gebhart, Aslin, & Newport, 2009; Weiss, Gerfen, & Mitchel, 2009). For example, *bi* would be the first syllable of *bidaku* in one stream but the final syllable of *labubi* in the other stream. Across several studies, participants have typically only learned units from the first familiarization stream. Bulgarelli and Weiss (2016) explored the possibility that this is due to overlearning of the first stream. To control for overlearning, Bulgarelli and Weiss split the familiarization for each stream into five 1-minute blocks, and participants were tested after each one. When participants were required to complete all five blocks of each stream regardless of test performance, only the first stream was learned. In contrast, when participants were transitioned

 doi:10.20944/preprints202005.0230.v1

to the second stream after reaching an accuracy criterion, they also appeared to learn that second stream. Thus, the first of two statistical streams is more readily learnable, although the effect may be mitigated by controlling for overlearning. Regardless, order plays a key role in what will be learned.

Order Sometimes Facilitates Learning

Although order can inhibit learning, it sometimes facilitates it. Carvalho and Goldstone (2017) find that order facilitates learning, although they emphasize that order modulates which aspects of the input a learner will attend to. The authors compared two familiarization orders. In blocked familiarization, category members were presented together, and different categories were presented in succession. In interleaved familiarization, members of different categories were interleaved throughout the familiarization. Based on a series of category learning experiments, Carvalho and Goldstone argue that blocked familiarization draws a learner's attention to within-category similarities, whereas interleaved familiarization leads learners to attend to the features that distinguish categories. Thus, different orders lead to differences in attention. What learners attend to will then determine what they learn.

Qian and Aslin (2014) note that order often has intuitive value, for example, from the perspective of a learner who infers seasons from changes in the weather over time. As a learner observes temperature changes, they come to understand that the seasons of winter, spring, summer, and fall represent a consistent, predictable cycle. Predicting seasons suggests how learners may use order to form clusters or categories of data, and using that logic, Qian and Aslin developed a computational model of order-based cluster learning. Their model accounts for a number of patterns in human cognition, including the perception of shooting streaks in basketball

 doi:10.20944/preprints202005.0230.v1

and the number of semantic topics contained within a conversation. For at least some learning problems, then, order is essential to what is learned, and humans are clearly attuned to it.

The Present Study

The literature reviewed above indicates that order is a powerful determinant of learning, although it has the capacity to both hinder and help. It is also worth noting that order is an inherent quality of language because any language input comprises elements arranged in a temporal sequence. It follows that order is endemic to statistical learning, as well. As Qian and Aslin (2014) point out, however, psychologists, linguists, and others typically attempt to nullify the effect of order through the randomization of stimuli. Although this is reasonable when order is not the focus of an experiment, there is still much to be learned about what order *does do* to the statistical learning process.

The present study is concerned with the effects of order on statistical learning, so order was explicitly controlled and manipulated. Participants learned the relative frequencies of item-medial consonant sequences during a familiarization. They subsequently rated test items containing the same sequences. The order of sequences was controlled during both familiarization and test. The word "order" here contrasts with randomization and refers to conditions in which items that share an item-medial consonant sequence appeared in succession. Consider a learner who is being familiarized with the consonant sequences /kt/, /mp/, and /st/, and where /st/ is high frequency in the familiarization relative to /kt/ and /mp/. An example of a randomized familiarization list might look like this: *gɪstək*, *nʌmpət*, *baɪstəm*, *nastək*, *paɪktəm*. In an ordered list, however, the three items containing /st/ all appear in succession, as in this

 doi:10.20944/preprints202005.0230.v1

example: *gɪstək*, *baɪstəm*, *nastək*, *nʌmpət*, *paɪktəm*. Thus, "order" here is specific to the order of word-medial consonant sequences across stimuli.

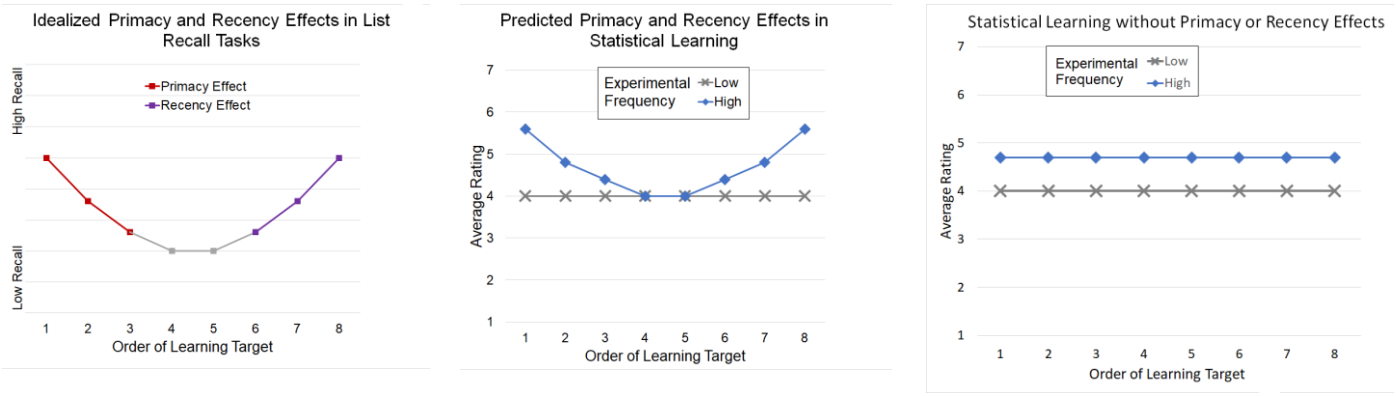
The present study is a comprehensive examination of order and randomization in the context of statistical learning of phonotactic frequencies by adults. It compares five different conditions in which randomization and order were applied separately to the familiarization and test blocks. Specific predictions related to the different conditions were not made, primarily because the operationalization of order does not closely follow previous studies. It is possible that an ordered list like *gɪstək*, *baɪstəm*, *nastək*, *nʌmpət*, *paɪktəm* could lead to local generalizations similar to those reported by Gerken and Quam (2017). However, it is just as possible that ordered lists could support cluster-based learning similar to that reported by Qian and Aslin (2014). Thus, this study takes an exploratory approach. We begin with an expectation based on a study with a similar design, Richtsmeier (2011), that participants should learn the relative frequencies of item-medial consonant sequences. In that study, the order of both familiarization and test items was always randomized. Higher ratings were given to test words containing frequent sequences relative to infrequent sequences. This relative difference in ratings was taken as evidence of learning. The question here is whether the same difference in ratings will be affected when the familiarization, the test, or both are ordered with respect to the targets.

Although the main analysis is exploratory, inclusion of order does allow for one specific hypothesis related to working memory. This is because statistical learning appears to depend on memory: For familiarization statistics to be tracked and learned from, the input must be stored in memory. Working memory is a sensible storage mechanism to posit for statistical learning, and in fact, there is evidence that statistical learning is constrained by working memory (Palmer & Mattys, 2016).

 doi:10.20944/preprints202005.0230.v1

Many studies show that working memory is sensitive to order (Howes, 2006), and there is therefore some reason to expect that statistical learning will be similarly affected by an ordered input. There are two well-known order effects in the working memory literature, typically referred to as primacy and recency effects. They are common in recall tasks where participants are asked to remember a list of words, such as the following list of 8 English words: *chair*, *power*, *woman*, *plane*, *flag*, *cost*, *seat*, *dream*. This list will exceed the average person's memory capacity, and only some of the words will be recalled (Cowan, 1998). The primacy effect refers to the finding that participants will typically recall more of the first items on the list, such as *chair* and *power*. The recency effect refers to better recall for the last items on the list, such as *seat* and *dream*. Figure 1a presents idealized primacy and recency effects; empirically observed primacy and recency effects depend on a variety of factors, including whether participants may recall the items feely or must recall them in order (for example, Jahnke, 1963; Raffel, 1936).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1



a. b. c.

Figure 1 – Figure 1a presents primacy and recency effects in immediate recall tasks based on a list with eight items. Figure 1b presents a hypothesis about how serial position effects might map onto a statistical learning task in which participants learn the relative frequencies of eight consonant sequences presented in order. Primacy and recency effects reflect the difference between ratings for high and low frequency sequences. Figure 1c presents statistical learning without serial position effects.

Assuming that statistical learning relies on working memory, if a statistical learning input is ordered so that items sharing a target appear in succession, it is possible that statistical learning could also reveal primacy and recency effects. Specifically, we may observe that a high experimental frequency advantage is strongest for target consonant sequences at the beginning of an ordered familiarization list, equivalent to a primacy effect, or to sequences at the end of the list, equivalent to a recency effect. Put differently, if a series of learning targets are presented in an ordered list, we might expect greater learning of the first and last targets in the series. This prediction corresponds with hypothetical results presented in Figure 1b. Note that this prediction would only hold for ordered data; randomized data does not allow for learning effects to be

 doi:10.20944/preprints202005.0230.v1

associated with specific positions in a list. Additionally, learning here is indexed solely by the difference between high frequency and low frequency conditions.

The prediction above assumes that statistical learning will exhibit specific working memory effects. Although memory is essential for successful statistical learning, the relevant memory store is poorly understood, and it is therefore unclear whether statistical learning should exhibit primacy and recency effects, even when the input is ordered. A lack of primacy and recency effects would correspond with hypothetical results that are presented in Figure 1c.

To summarize, the purpose of the present study is to explore the effects of order in a statistical learning task. By comparing ordered and randomized conditions, the results may reveal whether learning is sensitive to the difference between ordered and randomized inputs. Furthermore, the study probes for a specific type of working memory effect that depends on an ordered input. Because order is not well understood in statistical learning tasks, it is also possible that the results will reveal the importance of order in unanticipated ways.

Method

Participants

A total of 352 adult participants were recruited for the study. Participants were enrolled in psychology courses and received partial course credit as compensation. Participants' data were removed from analyses if participants were non-native speakers of English (n = 18); if they were multilingual or had more than two years of college-level training in a foreign language (n = 54); if they reported a history of language, hearing, or cognitive difficulties (n = 18); or if they met criteria for potential inattentiveness. Regarding inattention, participants were excluded if they gave the same rating 5 or more times in a row (n = 9), or if 4 or more of their ratings were made

 doi:10.20944/preprints202005.0230.v1

in under 300 ms (n = 2). During data collection, a design error was noted in some versions of the experiment. Data from these participants (n = 49) were not included in further analyses. Additionally, some data were lost or determined to be invalid due to experimenter error (n = 7). Data from the remaining 196 participants were entered into the statistical analyses.

Materials

Table 1 provides a subset of the materials used for this study (5 lists, 1 list per ordering of the 30 lists used in this study; see Appendix A for additional lists). Learning targets were the eight item-medial consonant sequences used by Richtsmeier (2011): /fp/, /kt/, /mk/, /mp/, /pk/, /sp/, /st/, and /ʃp/. The targets varied systematically in terms of their frequency in English: /kt/, /mp/, /sp/, and /st/ were frequent in English and /fp/, /mk/, /pk/ and /ʃp/ were infrequent, per the Phonotactic Probability Calculator (Vitevitch & Luce, 2004, https://calculator.ku.edu/phonotactic/, see Appendix B for a detailed report of the frequency calculations for experimental stimuli). In many studies, items containing high English frequency sequences are rated as more wordlike by native English speakers (for example, Albright, 2009). However, Richtsmeier (2011, 2016) has reported that high English frequency sequences can be rated lower if the instructions suggest that the items in which the sequences appear are from a language other than English.

Table 1 - A comparison of the five types of ordering of target item-medial consonant sequences. High English frequency targets are underlined. Targets with experimental frequency 3 appear in a bold font. Each ordering in this table represents one of the six lists per ordering that were used throughout the experiment.

	Ordering Conditions									
	Randomized Ordering		Randomized familiarization, ordered test		Ordered familiarization, randomized test		Ordered familiarization and test items (First-In-First-Out)		Ordered familiarization and test items (Last-In-First-Out)	
	n = 36		n = 39		n = 41		n = 38		n = 42	
	Target Order	Item Order	Target Order	Item Order	Target Order	Item Order	Target Order	Item Order	Target Order	Item Order
		fʌʃpət		fʌʃpət		fʌʃpət		fʌʃpət		fʌʃpət
		gɪfpək		gɪfpək	**ʃp**	daʃpək	**ʃp**	daʃpək	**ʃp**	daʃpək
		kɛʃpəs		kɛʃpəs		kɛʃpəs		kɛʃpəs		kɛʃpəs
		faspət		faspət	sp	faspət	sp	faspət	sp	faspət
		mɛstəm		mɛstəm	st	mɛstəm	st	mɛstəm	st	mɛstəm
		saʊktəs		saʊktəs		saʊktəs		saʊktəs		saʊktəs
		ʃampən		ʃampən	**kt**	paɪktəm	**kt**	paɪktəm	**kt**	paɪktəm
Familiariz-ation Block	*rando-mized*	zeɪktəs	*rando-mized*	zeɪktəs		zeɪktəs		zeɪktəs		zeɪktəs
		bapkəs		bapkəs	pk	bapkəs	pk	bapkəs	pk	bapkəs
		sæmkəsh		sæmkəsh	mk	sæmkəsh	mk	sæmkəsh	mk	sæmkəsh
		baɪfpəm		baɪfpəm		baɪfpəm		baɪfpəm		baɪfpəm
		daʃpək		daʃpək	**fp**	gɪfpək	**fp**	gɪfpək	**fp**	gɪfpək
		nafpək		nafpək		nafpək		nafpək		nafpək
		gumpən		gumpən		gumpən		gumpən		gumpən
		paɪktəm		paɪktəm	**mp**	ʃampən	**mp**	ʃampən	**mp**	ʃampən
		numpət		numpət		nʌmpət		nʌmpət		nʌmpət

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1

Table 1, continued

Test Block	*rando-mized*	biktəm nʌʃpək sæmpəf tuspən beɪmkəf mæfpəm koɪmpət foktəs kustəs lɛpkəf neɪstən poɪfpən zaʃpəm dɪmkəs tupkən lɪspət	**ʃp** sp st **kt** pk mk **fp** **mp**	zaʃpəm nʌʃpək tuspən lɪspət neɪstən kustəs biktəm foktəs lɛpkəf tupkən dɪmkəs beɪmkəf mæfpəm poɪfpən koɪmpət sæmpəf	*rando-mized*	biktəm nʌʃpək sæmpəf tuspən beɪmkəf mæfpəm koɪmpət foktəs kustəs lɛpkəf neɪstən poɪfpən zaʃpəm dɪmkəs tupkən lɪspət	**ʃp** sp st **kt** pk mk **fp** **mp**	zaʃpəm nʌʃpək tuspən lɪspət neɪstən kustəs biktəm foktəs lɛpkəf tupkən dɪmkəs beɪmkəf mæfpəm poɪfpən koɪmpət sæmpəf	**mp** **fp** mk pk **kt** st sp **ʃp**	sæmpəf koɪmpət poɪfpən mæfpəm beɪmkəf dɪmkəs tupkən lɛpkəf foktəs biktəm kustəs neɪstən lɪspət tuspən nʌʃpək zaʃpəm

The targets appeared medially in nonwords (referred to hereafter to as items) with a CVCCVC shape and with stress on the initial CVC syllable. Although there is some evidence that item-medial consonant sequences are more difficult to learn than word-initial and word-final consonants (Endress & Mehler, 2010), other studies have confirmed that participants are sensitive to the experimental frequencies of these sequences (Richtsmeier, 2011, 2016; Richtsmeier & Goffman, 2017). Recordings of the items for this study were culled from those latter studies.[1] All items were produced by adult female speakers of American English and were recorded in a sound booth. To ensure that participants had clear acoustic cues for each medial consonant, items were produced such that the first consonant of the second syllable was an onset. This was most relevant for items where the second syllable began with a stop consonant; those stops were produced with aspiration. All items were scaled to a standard 70 dB value of peak intensity.

 doi:10.20944/preprints202005.0230.v1

During familiarization, participants heard four of the target sequences in only one item and the other four targets in three items, a factor referred to as experimental frequency. Familiarization items were paired with make-believe animals as referents (Ohala, 1999). Familiarization lists always contained 16 items, with just one token per item. We anticipated higher ratings for experimental frequency 3 compared to 1 (Richtsmeier, 2011). “Higher ratings” are relative because a frequency of 3 is more relative to 1. Furthermore, there is no baseline, no absolute rating value, and no wrong answer that anchors the ratings. Thus, a learning effect is only observable based on the difference between experimental frequencies 3 and 1, but not based on the average rating given. Additionally, the familiarization is much briefer than those reported in similar statistical learning studies with adults (Denby et al., 2018; Finley, 2015; Onishi et al., 2002; Richtsmeier, 2011, 2016), so the effect of experimental frequency here was anticipated to be relatively small.

Participants heard items from 7-10 talkers, with the number of talkers depending on the list, and with 13 different talkers being used across all lists. In statistical learning research, talker variability is often beneficial (Bulgarelli & Weiss, 2018; Plante, Bahl, Vance, & Gerken, 2011; Richtsmeier, Gerken, Goffman, & Hogan, 2009; Richtsmeier & Goffman, 2017). However, the benefits are small, and talkers beyond the first two or three do not appear to provide an additional benefit (Richtsmeier & Moore, In preparation). The assignment of talkers to familiarization items in this experiment was not carefully controlled. Some talkers were used more than once per list, and some talkers were left out of a list if no good token of an item from that talker could be found. It was not anticipated that the subset of familiarization talkers or the number of talkers would influence the results, however.

 doi:10.20944/preprints202005.0230.v1

Separate items were used for test, although those items contained the same eight item-medial consonant sequences. At test, participants provided ratings for two items containing each of the eight targets, or 16 total test items. Two talkers provided productions of the test items, with just one of the two talkers used for a given list. Additionally, two sets of test items were alternated across the different lists, or 32 unique test items overall. Because the test items only appeared at test, an experimental frequency effect would indicate that participants had learned the frequencies of the target sequences rather than specific familiarization items.

The order in which familiarization and test items were presented was manipulated systematically, a factor referred to as ordering. Table 1 above provides examples of the five different ordering conditions as well as the number of participants from each ordering who were entered into statistical analyses. In the randomized ordering, both familiarization and test items appeared in a random order uniquely generated for each participant by Paradigm experimental software. The randomized ordering reflects common practice in much of statistical learning research (Denby et al., 2018; Finley, 2015; Gerken & Quam, 2017; Onishi et al., 2002; Richtsmeier, 2011, 2016, etc).[2] In the randomized familiarization, ordered test condition (hereafter ordered test), familiarization items appeared in random order, but the test items were ordered by their medial consonant sequences. In the ordered familiarization, randomized test condition (hereafter ordered familiarization), the familiarization items were ordered, but the test items were randomized. Both familiarization and test items were ordered in the first-in-first-out and last-in-first-out orderings, and as their names suggest, the two conditions differed in terms of the relative order of familiarization and test. In the first-in-first-out ordering, test order recapitulated familiarization order. In the last-in-first-out ordering, the final target from the familiarization was the first target in the test. In sum, the five ordering conditions provide

 doi:10.20944/preprints202005.0230.v1

comprehensive coverage of the possible combinations of order and randomization for both familiarization and test within this experimental design.

Steps were taken to counterbalance the experimental variables and to limit the effects of nuisance variables. For each of the five orderings, six different lists were generated, or 30 total lists (see Appendix A). Across lists, the order of the targets varied, and so did the eight targets' experimental frequencies. For example, although Table 1 shows /ʃp, kt, fp, mp/ as high experimental frequency, they were low experimental frequency in three of six lists. As a result of counterbalancing, each level of English frequency (high and low) and each level of experimental frequency (3 and 1) occurred in each of the eight target positions possible for the non-random orderings.[3] However, not every consonant sequence appeared in each of the eight positions. This and other limitations of the design are reviewed in the General Discussion.

Procedure

The procedure largely follows that reported by Richtsmeier (2011). At the start of the experiment, participants were told that they would be learning the names of make-believe animals in a "made-up" language. During the familiarization, the task was to watch the make-believe animals and listen to their names. At test, participants heard new items that were described as potential make-believe animal names. Their task was to rate each test item on a 1-7 Likert scale, where 1 meant "not a potential animal name", 4 meant "neutral", and 7 meant "a great potential animal name". Following the rating task, participants completed a real-word list memory task and a nonword repetition task, the results of which will be reported elsewhere. Once all tasks were complete, the experimenter asked participants, "In the first part of the experiment, where you listened to and then rated make-believe animal names, what do you think

 doi:10.20944/preprints202005.0230.v1

the purpose of that part of the experiment was?" The reason for asking this question was to ensure that participants engaged in implicit learning of the statistical manipulations. This probe was implemented after the experiment had begun, however, and only 273 of the 352 recruited participants completed it. However, no participant who completed the probe reported having been aware of the frequency of item-medial consonant sequences. The most common response given was, "I don't know," (n = 140).

Analysis

Ratings were scanned according to pre-established criteria to filter out results from potentially inattentive participants. As described above, data were excluded if the participants made the same rating five or more times in a row, or if they made four or more ratings in less than 300 ms.

Main statistical analysis. This section describes the baseline statistical model developed for the main analysis. The purpose of this analysis was to determine whether the expected experimental frequency effect (3 > 1) might vary depending on how the stimuli were ordered. Overall, the experimental design involved a multitude of potential cells, including two levels of English frequency, two levels of experimental frequency, eight levels of target order,[4] and five levels of ordering. Most of the possible interactions of a full factorial model were not of interest, so an exploratory mixed model analysis was implemented using guidelines suggested by Baayen, Vasishth, Kliegl, and Bates (2017). The initial model for the main analysis included main effects for English frequency, experimental frequency, ordering, and target order, as well as three random effects, described below. Inclusion of English and experimental frequencies is supported by previous research (for example, Albright, 2009; Bailey & Hahn, 2001; Onishi et al., 2002;

 doi:10.20944/preprints202005.0230.v1

Richtsmeier, 2011, 2016) and therefore sensible in a baseline model. Main effects of ordering and target order were also included to allow for a straightforward comparison with the alternative models that included those factors in interactions. Because the primary goal of the main analysis was to determine whether learning would be affected by ordering, the first alternative model allowed for an interaction of experimental frequency and ordering. Additional interactions would be added assuming a better fit to the data. All interactions would then be broken down to better understand whether an experimental frequency effect was enhanced or attenuated in one or more of the ordered conditions.

Baayen et al. (2017) emphasize how experimental data in a time series often contain a human element: Participants' attention may wander, they may develop and implement unique strategies for responding, and they may have stronger memories for some items compared to others. All of these aspects of cognition can result in autocorrelation, or a frequently-nonlinear mathematical dependency among the various dependent measures. Generalized additive mixed modeling with random effects is one way to account for these nonlinear or "wiggly random effects" in the statistical analyses, and Baayen and Divjak (2017) show how this type of model can be applied to wordlikeness ratings. To address autocorrelation in generalized additive mixed modeling, Baayen et al. suggest that analyses begin with data visualization to determine whether autocorrelation exists, and whether wiggly random effects are needed.

Figure 2 below provides individual data plots for the first five participants from each of the five orderings. At least three by-participant trends appear to be possible in these data, suggesting that autocorrelation is indeed present and should be accounted for. Here, a series of random effects are described that may account for autocorrelation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1

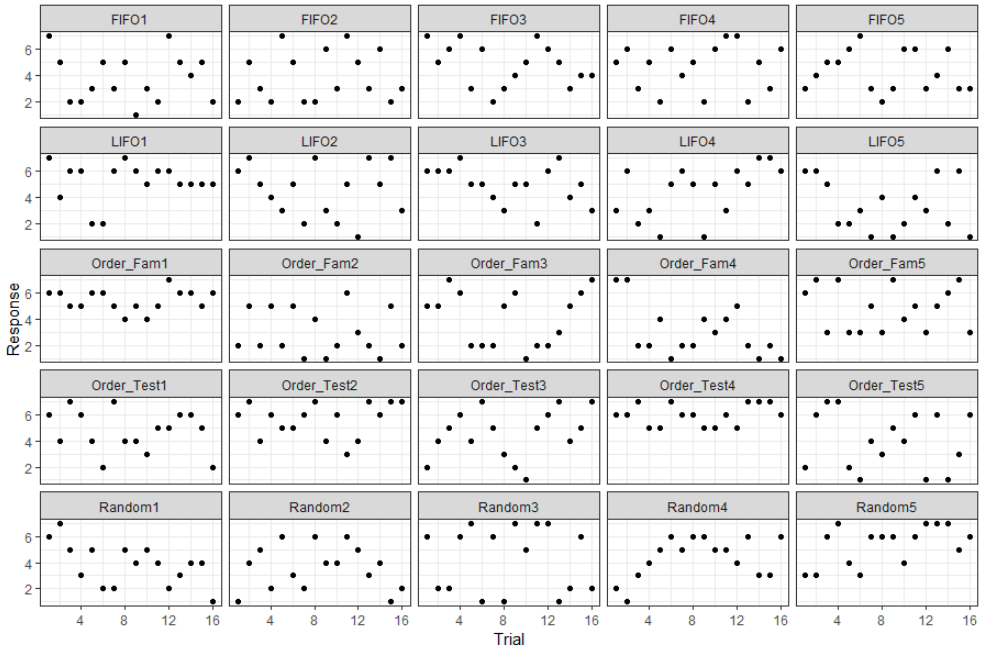


Figure 2 – Plots of individual responses across trials for the first five participants completing each ordering.

First, participants differed in terms of their first rating, which could be modeled with a subject-level random effect for intercepts. Initial model testing determined that by-subject intercepts were not a significant predictor, however, so this factor was not included in the baseline model or subsequent models. Second, some participants used more of the 1-7 rating scale than others, a difference that can be captured with a random effect for the participants' standard deviations. Third, following Baayen et al. (2017), time-based correlations—such as pegging a rating to the previous rating—appeared to be present across the participants' ratings. These were also modeled with wiggly random effects.

 doi:10.20944/preprints202005.0230.v1

Wiggly random effects are modeled mathematically using regression splines, or composite equations composed of simpler basis functions. Basis functions account for the curves or wiggles in the data. Two types of wiggly random effects are relevant for the analyses conducted here: thin plate regression splines and tensor product smooths. Thin plate regression splines are the more basic of the two, and they are appropriate for continuous univariate predictors. Participants' standard deviations were modeled using a thin plate regression spline. Tensor product smooths are appropriate when multiple predictors are combined and the two predictors have different ranges of possible values. A tensor product smooth was used to model the interaction of participants and experimental time because participants had a range of 196 possible values (one value per participant), whereas experimental time could be modeled either by the number of trials or by the number of target consonant sequences. Fitting ratings to the 8 positions of the target sequences provided a better fit for the data than the 16 values of trial, so target order was used in the interaction of participants and experimental time. We return to this choice in the General Discussion.

A wiggly random effect was also added for the 32 different test items. Initial model testing determined that the sum of phone frequencies reported by the Phonotactic Probability Calculator (Vitevitch & Luce, 2004)—fitted with a thin plate regression spline—provided a better fit for item variation compared to the sum of biphone frequencies and compared to the inclusion of both sums. One interesting consequence of modeling phonotactic probabilities using wiggly random effects is that it allows the statistical model to capture nonlinear responses to different phonotactic frequencies. Responses to phonotactic frequencies might be expected to be linear, with each increase in a target sequence's probability resulting in an equivalent change to participant ratings. However, Baayen et al. (2017) show that a nonlinear function often connects

 doi:10.20944/preprints202005.0230.v1

a word's frequency to corresponding lexical decision times; more specifically, some participants in their study were particularly sensitive to the mid-range of word frequencies compared to the high and low word frequencies. Here, a significant wiggly effect of the sum of phone frequencies would suggest a nonlinear function best maps wordlikeness ratings to phonotactic probabilities.

In summary, three wiggly random effects were included in the main analysis: by-participant standard deviations, the interaction of participant by experimental time, and by-item sums of phone probabilities for each test item. For each wiggly random effect, five basis functions were used. More basis functions correspond to more wiggles, however, modifications to this default value of five did not improve model fit. Additionally, the proportionality constant ρ was not adjusted in the reported models. Increasing ρ typically removes autocorrelation, but adjustments to ρ also did not improve model fit.

The mixed models for the main analysis were implemented using the *gam* function in the *mgcv* package in R. Model comparisons were made using a χ^2 test of the models' effective degrees of freedom and restricted maximum likelihood (REML) using the *itsadug* package. Further discussion of many of these choices can be found in Baayen and Divjak (2017) and Baayen et al. (2017).

Statistical analysis based on predictions derived from working memory research. The second analysis was motivated by working memory research. The working memory literature suggests that items at the beginning and end of a list are better remembered, and there is some evidence that working memory constrains statistical learning (Palmer & Mattys, 2016). Based on that logic, the hypothesis is that experimental frequency effects will be isolated to the beginning and end of a statistical learning task in which the familiarization is ordered, that is, to

 doi:10.20944/preprints202005.0230.v1

the first and/or last targets in the first-in-first-out and last-in-first-out conditions. The alternative hypothesis is that learning will be consistent across target order. To compare these hypotheses, the baseline model described above was fitted using only data from the first-in-first-out and last-in-first-out conditions. A planned alternative model allowed experimental frequency to interact with target order. A third planned alternative model allowed experimental frequency to interact with a quadratic term for target order, as would be expected based on the shape of primacy and recency effects (see Figure 1).

Results

Results of the main analysis

Figure 3 presents the ratings made across the eight word-medial target positions, with separate plots for each of the five ordering conditions. The raw ratings are overlaid by Loess curve estimates of the average ratings and confidence intervals over the eight positions of target order for both experimental frequency 1 and 3. The average rating across all conditions was 4.36.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1

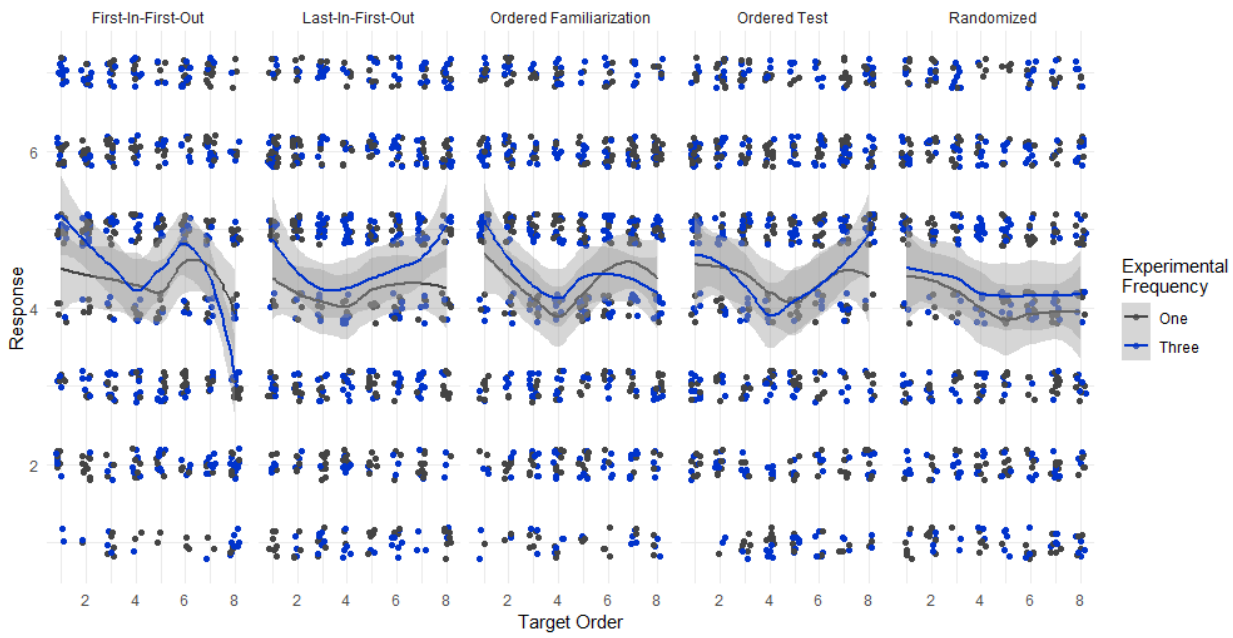


Figure 3 – A plot of ratings by target order and ordering condition. Average ratings and confidence intervals across target order were plotted using Loess curves.

The baseline model—including the intercept, the four main effects (English frequency, experimental frequency, target order, ordering) and three random effects (participants' standard deviations, sums of phone probabilities of the test items, and participant-by-target-order interactions)—is summarized in Table 2. Regarding parametric effects, there was a main effect of experimental frequency (p = 0.018) as participants gave higher ratings to test items with experimental frequency 3 (M = 4.43) compared to test items with experimental frequency 1 (M = 4.29). There was also a main effect of target order (p < 0.001, slope = -0.15), suggesting that participants gave increasingly lower ratings for each subsequent target consonant sequence. Ratings did not differ significantly based on English frequency or the ordering condition. All three smooth terms were significant and are described in greater detail below.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1

Table 2 – Summary of the baseline generalized additive mixed model. Target order was scaled to be symmetrical around the intercept. A smooth term marked as s() denotes a thin plate regression spline, te() a tensor product smooth. The ordered test condition was used as the comparison group for the analysis of ordering.

Parametric coefficients	Estimate	Standard Error	*z*-value	*p*-value
Intercept	5.68	0.91	6.27	<0.001*
English Frequency	-0.01	0.08	-0.13	0.894
Experimental Frequency	0.15	0.06	2.38	0.018*
Target Order	-0.15	0.04	-3.45	<0.001*
Ordering = randomized	-0.22	0.16	-1.36	0.174
Ordering = ordered familiarization	-0.05	0.11	-0.43	0.671
Ordering = first-in-first-out	0.07	0.12	0.55	0.583
Ordering = last-in-first-out	-0.05	0.11	-0.38	0.703
Smooth Terms/Random Parameters	**Effective degrees of freedom**	**Ref. degrees of freedom**	**X^2**	***p*-value**
s(standard deviation)	0.99	1.00	76.40	<0.001*
s(Phone Sum)	0.92	1.00	11.17	<0.001*
te(Participant × Target Order)	12.65	15.81	44.67	<0.001*
Model Summary				
Deviance explained	1.45%			
REML	5915.3			

To determine whether the experimental frequency effect depended on ordering, the baseline model was compared to a model in which experimental frequency and ordering

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1

interacted. Table 3 provides the outcome of the χ^2 test. The baseline model had a lower REML score, and the comparison was not significant. Thus, the exploratory statistics did not advance beyond this first comparison, and the baseline model was accepted as optimal. Accepting the baseline model—in which ordering was not significant—suggests that ordering and randomization do not differentially impact how participants interpret the experimental frequency manipulation.

Table 3 – A comparison of the baseline model to an alternative model in which experimental frequency and ordering were allowed to interact.

Model	REML Score	Effective degrees of freedom	Difference	df	*p*-value
Baseline model	5915.34	15			
Model with Experimental Frequency x Ordering interaction	5917.37	19	2.03	4	n.s.

The three wiggly random effects from the main analysis are now described in greater detail. The first effect was the interaction of participants and standard deviations ($p < 0.001$). A graph of average ratings by standard deviations appears in Figure 4 below. Participants with greater standard deviations generally had lower average ratings, although average ratings plateaued and started to rise again for standard deviations near or above 2.0. This finding suggests that participants who used less of the rating scale—resulting in lower standard

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1

deviations—tended to give relatively high ratings on average, although it is not clear why this pattern would hold.

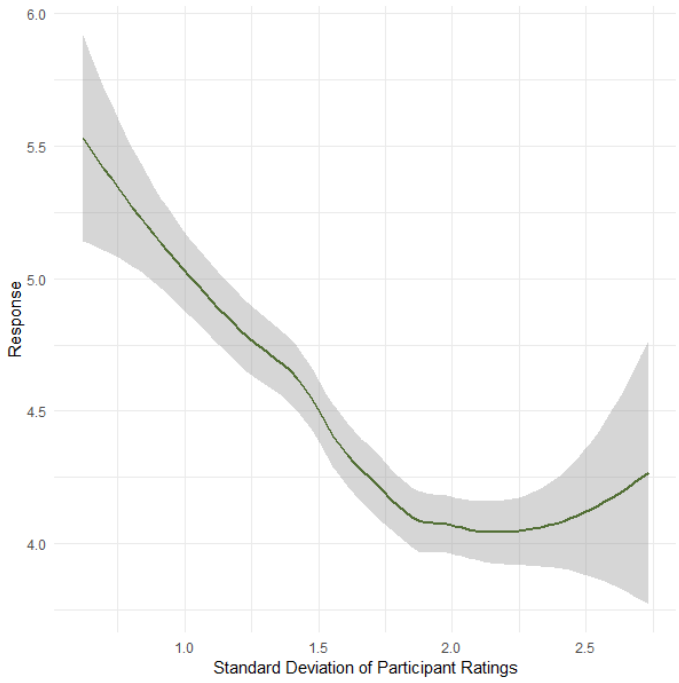


Figure 4 - Average responses plotted by the participants' standard deviations.

The second random effect was the interaction of participants and experimental time ($p < 0.001$), the latter being modeled with target order. This result suggests participants differed in terms of how they made their ratings over time. Although individual differences are to be expected in any cognitive task, the significant wiggly random effect suggests that participants were responding strategically, and their strategies can be modeled using regression splines. To flesh out this result, artificial data were created. The artificial data were generated using separate frequency bins for each rating (1-7) separately for experimental frequencies 1 and 3. The data were then randomly assigned to trials (1-16) and the five ordering conditions. Thus, the artificial

 doi:10.20944/preprints202005.0230.v1

data recapitulated the main effect of experimental frequency but were otherwise randomly generated. When the artificial data were entered into the main analysis, there was an expected effect of experimental frequency, but the random effect for the interaction of participants and target order was not significant ($p = 0.562$). The interaction of participants and experimental time in the human data tells us that participants impose order on their own responses. Participant-imposed orders are idiosyncratic but structured, a fact which is captured mathematically by the regression splines.

The third wiggly random effect was also significant ($p < 0.001$). This effect captured by-item differences in responding related to the sums of the phone probabilities of the test items. Differences in ratings based on the sums of the phone probabilities, with separate lines for high and low English frequency items, appear in Figure 5 below. Overall ratings for the high and low English frequency sequences appear as dashed lines. Although not significant, the numerically higher ratings for low English frequency sequences replicates findings reported by Richtsmeier (2011, 2016) and appears to be the result of directions in which participants are asked to consider the items as not coming from English. More striking than the average difference between high and low English frequency sequences, however, are the strongly nonlinear effects of the phone sums. Participants appeared to give the highest ratings for the low English frequency test items with phone sums of about 1.25, but ratings did not consistently drop as the phone sums increased. Although it is not clear what caused these nonlinear effects, the nonlinear relationship between phone sums and wordlikeness ratings is reminiscent of the nonlinear relationship between word frequency and reaction times discussed by Baayen et al. (2017). Future research is needed to better detail when human knowledge of frequency is not a direct reflection of the statistics.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1

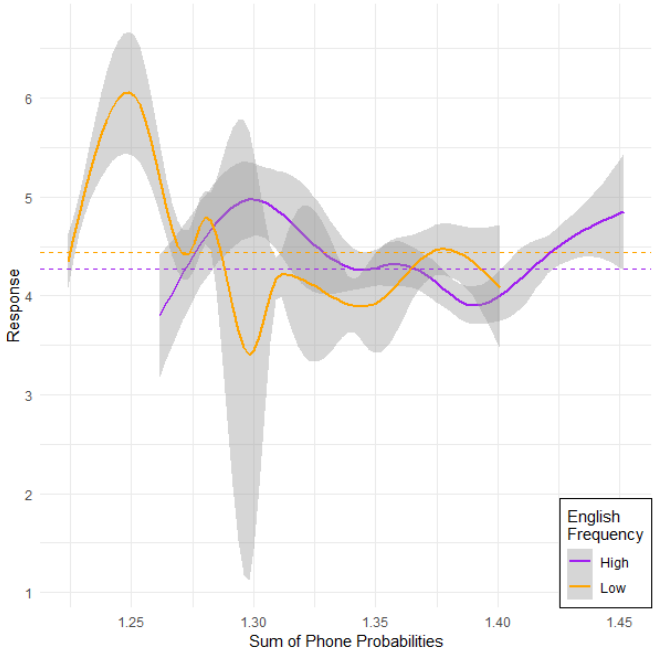


Figure 5 – Average responses plotted by the sum of phone probabilities of the test items. High and low English frequency medial consonant sequences are plotted separately.

Results of the analysis based on predictions derived from working memory

To test whether ordered familiarizations would result in primacy and recency effects, the baseline model was run using only data from the first-in-first-out and last-in-first-out orderings (see Appendix B for a summary of this model). This baseline model was then compared to two alternative models in which target order interacted first with a linear term and second with a quadratic term for experimental frequency. Table 4 provides the outcome of the χ^2 tests—neither comparison was significant. Thus, the results did not support the hypothesis that working-memory-based primacy and recency effects occur in statistical learning.

 doi:10.20944/preprints202005.0230.v1

Table 4 – A comparison of the baseline model (limited to data from the first-in-first-out and last-in-first-out orderings) to an alternative model in which experimental frequency and target order were allowed to interact.

Model	REML Score	Effective degrees of freedom	Difference from Baseline model	df	*p*-value
Baseline model with first-in-first-out and last-in-first-out data only	2410.74	12			
Model with Experimental Frequency x linear Target Order interaction	2410.60	13	0.14	1	0.599
Model with Experimental Frequency x quadratic Target Order interaction	2415.24	13	4.50	1	n.s.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2020 doi:10.20944/preprints202005.0230.v1

Figure 6 plots the difference between experimental frequencies 3 and 1 across the 8 positions of target order, with different lines for each ordering. Difference scores for the first-in-first-out and last-in-first-out conditions are bolded. Figure 6 indicates that—although there was variation in the effect of experimental frequency—the magnitude of the effect did not vary systematically with either target order or ordering.

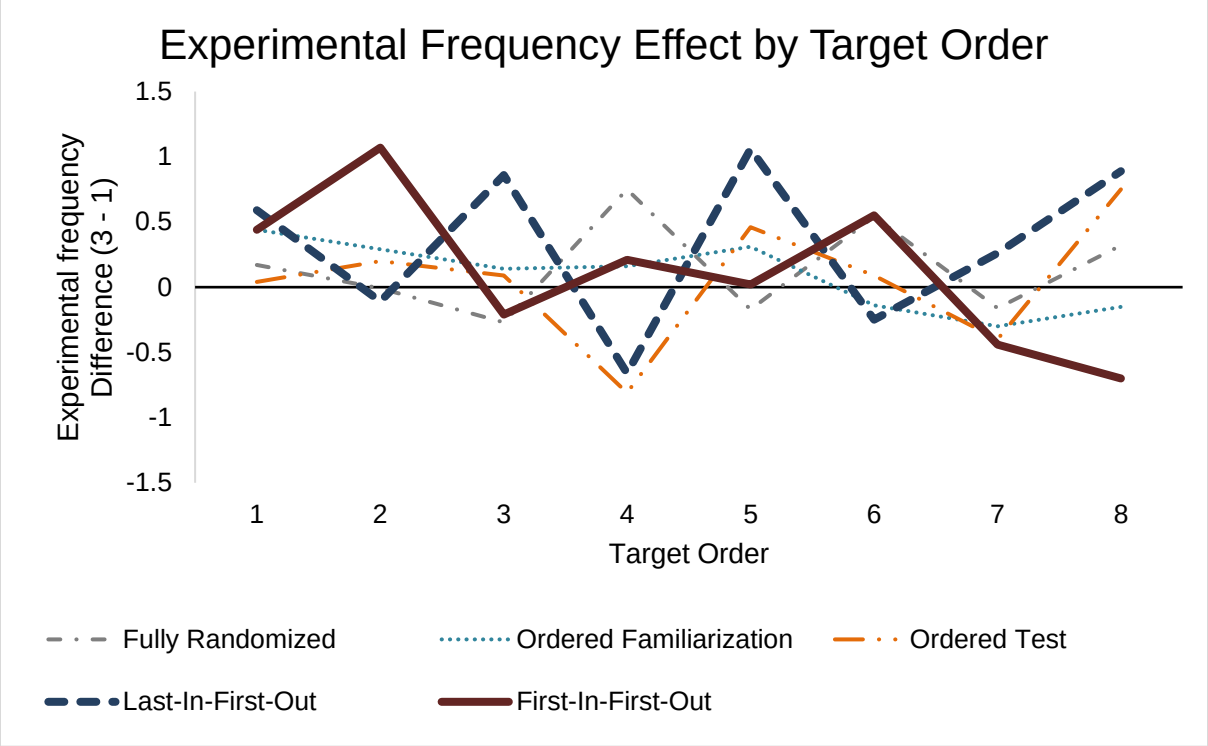


Figure 6 – Average differences between experimental frequencies 3 and 1 across target order and ordering. Values greater than 0 indicate higher ratings for experimental frequency 3 targets.

 doi:10.20944/preprints202005.0230.v1

General Discussion

In this study, we considered how statistical learning would be affected by the intentional ordering of stimuli. Participants completed a statistical learning task where item-medial consonant sequences served as the learning targets, and the experimental frequency of those targets varied. Overall, participants gave higher ratings to test items in the high experimental frequency familiarization. This finding is in line with other published statistical learning effects for phonotactics (Denby et al., 2018; Onishi et al., 2002; Richtsmeier, 2011).

The primary question raised in the introduction was whether the benefit of high experimental frequency would be modulated by ordering the familiarization phase, the test phase, or both. Five ordering conditions were compared. No significant differences were found among those conditions, nor did ordering interact significantly with experimental frequency. A reasonable conclusion based on the data is that the ordered input supported learning to the same degree as randomization. Furthermore, in contrast to expectations based on the working memory literature, it did not appear that the experimental frequency effect was strongest at the beginning or end of the ordered familiarization lists. In other words, although the design allowed for primacy and recency effects that are observable in list recall tasks, it does not appear that these effects exist in the context of this statistical learning task. Below, we return to the literature on order effects that was reviewed in the Introduction. The discussion addresses why an ordered input did not appear to either hinder or help learning. The lack of a working memory effect is also discussed, as are some weaknesses in the design and possible directions for future research.

 doi:10.20944/preprints202005.0230.v1

Does Order Help or Hinder Statistical Learning?

Previous research has shown that certain types of order can inhibit learning. For example, Gerken and Quam (2017) found that infants failed to learn a general phonological pattern in the input when salient local patterns were also present. Bulgarelli and Weiss (2016) and others have found that, when statistical input streams have conflicting statistics, participants will often only learn the first of two inputs.

Although the present design differs from those reported by Gerken and Quam (2017) and Bulgarelli and Weiss (2016), it is worth considering a few reasons why the ordered conditions here did not inhibit learning. One reason that local generalizations may not have been an issue is that adults have sufficient memory capacity to support multiple generalizations, that is, generalizations about each of the four high experimental frequency targets. In support of this interpretation, Gerken, Quam, and Goffman (2019) compared infant and adult performance on a statistical learning task. The patterns were “either-or”, for example, that both consonants in a CVCV word were either voiced or voiceless. Gerken et al. (2019) found that infants readily learned these either-or patterns, whereas the adults struggled. They interpret the infants’ advantage as indicative of their limited memories—an infant’s limited memory leads them to attend more to local generalizations, whereas the larger memory capacity of adults allows for more diffuse generalizations at the expense of local ones. The majority of adults in our study may have had sufficient memory capacity to allow for generalizations over multiple phonotactic sequences, even when those sequences occurred in serial order.

When the present results are compared to multistream statistical learning studies such as Bulgarelli and Weiss (2016), it is worth considering that the relative frequencies of medial

 doi:10.20944/preprints202005.0230.v1

sequences like /kt/, /mp/ and /st/ are not inherently in conflict in the same way as the transitional probabilities of the two inputs in Bulgarelli and Weiss' study. Thus, the first language in a multistream statistical learning task may be more learnable as much because of conflicting statistics as because of order. One way to test this possibility would be to run a multistream statistical learning study in which the syllables and words of the two inputs are entirely different. If conflict is essential to the inhibition of learning, participants in such a study may better learn the second input.

The introduction also reviewed cases where order supports or has been central to learning (Carvalho & Goldstone, 2017; Qian & Aslin, 2014). The present results are consistent with the claim that order can facilitate learning, although they do not support a claim that order is superior to a randomly-ordered input. Thus, when considering the benefits of order, it is important to consider why all the ordering conditions appeared to support learning.

Carvalho and Goldstone (2017) describe how learning is affected by blocked and interleaved familiarizations. Blocked familiarization appears to support learning of within-category structure, whereas interleaved familiarization supports learning of category differences. The ordered conditions here resemble blocked familiarization in some ways and interleaved familiarization in other ways. Here, the ordered conditions were blocked with respect to the eight learning targets, but they were interleaved with respect to experimental frequency. As can be seen in Appendix A, none of the six lists grouped together either the four experimental frequency 3 targets or the four experimental frequency 1 targets. From the perspective of interleaving, it may be that the ordered and randomized conditions were essentially equivalent. Perhaps the results were equivalent across the ordering conditions because of interleaving experimental

 doi:10.20944/preprints202005.0230.v1

frequencies. It is not possible to compare blocked and interleaved experimental frequency based the present study, but such a comparison may fruitful in future research.

Qian and Aslin (2014) report that human learners are highly attuned to order. They elaborate a computational model which forms clusters of data based on temporal proximity, that is, based on order. Although the ordered input was not superior to the randomized input in terms of supporting learning, the Qian and Aslin model may yet be helpful for explaining one of the results in which order did appear to matter. Recall that initial model testing was better able to fit the data when experimental time was modeled with target order rather than trial number. Target order may have been the better predictor because participants provided relatively coherent ratings for a given cluster. In other words, by using target order, variance in the model was reduced—at least for the conditions with an ordered test—because participants tended to provide similar ratings for the two words sharing a word-medial consonant cluster.

The Qian and Aslin (2014) model may also help explain the random wiggly effect for the interaction of participants and target order. This random effect indicates that participant responses over time were patterned, revealing idiosyncratic orders of responding imposed by the participants themselves. The human data contrasted with randomly generated data in which there was not an interaction of participant and target order. Consistent with Qian and Aslin (2014), this effect tells us that human participants respond in a way that is nonrandom, strategic, and dependent on order. One possible interpretation of this effect is that participants peg each rating to the rating they made previously. Each rating may depend on the last one, resulting in time-based dependencies. Such a dependency appears to be natural given that the materials are unfamiliar. Consider a participant who completed the randomized ordering in Table 1. That participant may struggle to decide whether /nʌʃpək/ is a good make-believe animal name in an

 doi:10.20944/preprints202005.0230.v1

absolute sense, but they may find it easier to say whether /nʌʃpək/ is better or worse than the previous item /biktəm/. It may be possible to confirm such a response strategy with random walk modeling (Fific, Little, & Nosofsky, 2010). Nevertheless, because ordering was not a significant predictor of the results, it does not appear that these participant-specific orders were driven by the order of the familiarization or test items.

Working Memory in Statistical Learning

In contrast to initial predictions based on working memory, there was no evidence that participants better learned targets that appeared at the beginning or end of the ordered familiarization lists. As such, a well-studied phenomenon from the working memory literature does not appear to arise in statistical learning. A sensible conclusion is there is still much to learn about the nature of the memory system that subtends statistical learning. Although primacy and recency effects were not observed here, working memory may play a role in statistical learning. For example, Palmer and Mattys (2016) used speech rate in a statistical segmentation task to establish a link to working memory. Across three experiments, participants learned to find sequences like *bidaku* in a syllable stream, and they were more successful when the speech rate was slowed. However, the benefits of a slow rate were attenuated by adding N-back shape- and rhyme-matching tasks. Palmer and Mattys note that working memory is commonly disrupted by secondary language-based tasks. Since statistical learning performance suffered a similar disruption, the authors conclude that working memory partially determines what can be learned (see also Noonan, 2014 for similar results).

For learners to be sensitive to any frequency manipulation, some memory system is needed to track the statistics of the input. Work by Palmer and Mattys (2016) suggests that

 doi:10.20944/preprints202005.0230.v1

working memory has some explanatory power, but when interpreting the effect of competing tasks on statistical learning, it will be important for future research to disentangle working memory and attention (Fernandes, Kolinsky, & Ventura, 2010). The lack of primacy and recency effects here further complicates our understanding of what memory system supports statistical learning. As a next step, we plan to explore working memory capacity in statistical learning by correlating by-participant experimental frequency effects to the results from the free word recall and nonword repetition tasks that our participants also completed. Working memory capacity may predict sensitivity to the experimental frequency manipulation.

Weaknesses of the Design

There are at least two weaknesses of the current study that should be acknowledged. First, the design included a large number of factors, but the complexity of the design was managed with a relatively small number of lists for counterbalancing. Some imbalances resulted, including that the eight target sequences did not appear in every position of target order, let alone as both high and low experimental frequency in every position of target order. To better understand the role of order in statistical learning, future studies may benefit from focusing on a smaller number of conditions, for example, solely on ordered conditions.

A second concern is that the lack of interactions between ordering and target order could reflect insufficient power. Even with data from nearly 200 participants, it would have been difficult to unpack all possible interactions. The present results may indicate that effects established in one area of cognition will not always translate to another area, but large sample sizes and replications are necessary to ensure that unobserved effects are truly absent, rather than lacking the necessary amount of supporting data.

 doi:10.20944/preprints202005.0230.v1

Conclusion

The present study replicates the finding that adults are able to learn relatively fine-grained phonotactic frequencies, in this case, the frequencies of word-medial consonant sequences. Participants learned these frequencies in a short amount of time, with relatively sparse data, and regardless of the order in which the sequences were presented. Still, participants were clearly influenced by order in that they appeared to order their own responses. Future studies of order in the context of statistical learning of phonotactic probabilities may help link this measure of statistical learning to memory and other areas of cognition.

 doi:10.20944/preprints202005.0230.v1

Acknowledgments

The author extends gratitude to Allison Brockette for help running participants. Improvements to this manuscript were made via feedback from LouAnn Gerken. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

 doi:10.20944/preprints202005.0230.v1

References

Adriaans, F. W., & Kager, R. (2017). Learning novel phonotactics from exposure to continuous speech. *Laboratory Phonology, 8*(1), 1-14.

Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology, 26*(1), 9-41.

Baayen, R. H., & Divjak, D. (2017). Ordinal GAMMs: a new window on human ratings. In *Thoughts on Language: Studies in Cognitive Linguistics in Honor of Laura A. Janda* (pp. 1-13). Bloomington, IN: Slavica Publishes.

Baayen, R. H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language, 94*, 206-234.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44*(4), 568-591.

Bulgarelli, F., Benitez, V., Saffran, J., Byers-Heinlein, K., & Weiss, D. J. (2017). *Statistical Learning of Multiple Structures by 8-Month-Old Infants.* Paper presented at the Proceedings of the... Annual Boston University Conference on Language Development. Boston University Conference on Language Development.

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.

Bulgarelli, F., & Weiss, D. J. (2018). *The More the Merrier? The Impact of Talker Variability on Artificial Grammar Learning in Preschoolers and Adults.* Paper presented at the Boston University Conference on Child Language Development.

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(11), 1699.

Cowan, N. (1998). Visual and auditory working memory capacity. *Trends in cognitive sciences, 2*(3), 77.

Denby, T., Schecter, J., Arn, S., Dimov, S., & Goldrick, M. (2018). Contextual variability and exemplar strength in phonotactic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(2), 280.

Endress, A. D., & Mehler, J. (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 235.

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics, 72*(6), 1522-1532.

Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review, 117*(2), 309.

Finley, S. (2015). Learning nonadjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language, 91*(1), 48.

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.

Gerken, L., & Quam, C. (2017). Infant learning is influenced by local spurious generalizations. *Developmental Science, 20*(3). doi:10.1111/desc.12410

Gerken, L., Quam, C., & Goffman, L. (2019). Adults Fail to Learn a Type of Linguistic Pattern that is Readily Learned by Infants. *Language Learning and Development*, 1-16.

Howes, M. B. (2006). *Human memory: Structures and images*: Sage Publications.

Jahnke, J. C. (1963). Serial position effects in immediate serial recall. *Journal of Verbal Learning and Verbal Behavior, 2*(3), 284-287.

Noonan, N., B. (2014). *The Relationship between Implicit and Explicit Processing in Statistical Language Learning.* The University of Western Ontario, Retrieved from https://ir.lib.uwo.ca/etd/2130 (2130)

 doi:10.20944/preprints202005.0230.v1

Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders, 32*(6), 397-422.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*(1), B13-B23.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology, 69*(12), 2390-2401.

Paradigm. (2015). Paradigm Experimental Software (Version 2.4) [Computer software]. Lawrence, KS: Perception Research Systems Incorporated. Retrieved from http://www.paradigmexperiments.com/

Pierrehumbert, J. B. (1994). Syllable structure and word structure: A study of triconsonantal clusters in English. In P. A. Keating (Ed.), *Papers in Laboratory Phonology III* (pp. 168-190). Cambridge, U.K.: Cambridge University Press.

Plante, E., Bahl, M., Vance, R., & Gerken, L. (2011). Beyond phonotactic frequency: presentation frequency effects word productions in specific language impairment. *Journal of Communication Disorders, 44*(1), 91-102. doi:10.1016/j.jcomdis.2010.07.005

Qian, T., & Aslin, R. N. (2014). Learning bundles of stimuli renders stimulus order as a cue, not a confound. *Proceedings of the National Academy of Sciences, 111*(40), 14400-14405.

Rácz, P., Hay, J., Needle, J., King, J., & Pierrehumbert, J. B. (2016). Gradient Maori phonotactics. *Reo, Te, 59*, 3.

Raffel, G. (1936). Two determinants of the effect of primacy. *The American Journal of Psychology, 48*(4), 654-657.

Richtsmeier, P. T. (2011). Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology, 2*(1), 157-183.

Richtsmeier, P. T. (2016). Phonological and Semantic Cues to Learning from Word-Types. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 7*(1).

 doi:10.20944/preprints202005.0230.v1

Richtsmeier, P. T., Gerken, L., Goffman, L., & Hogan, T. (2009). Statistical frequency in perception affects children's lexical production. *Cognition, 111*(3), 372-377. doi:10.1016/j.cognition.2009.02.009

Richtsmeier, P. T., & Goffman, L. (2017). Perceptual statistical learning over one week in child speech production. *Journal of Communication Disorders, 68*, 70-80. doi:10.1016/j.jcomdis.2017.06.004

Richtsmeier, P. T., & Moore, M. (In preparation). Children's speech accuracy: influences of production practice, input frequency, and talker variability.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 1926-1928.

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.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America, 119*(1), 597-607.

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.

 doi:10.20944/preprints202005.0230.v1

Footnotes

[1] As a result of culling the items from other experiments, across lists the targets appeared in a variable number of items. A given participant was always exposed to either one or three items sharing a target, however, and they always rated two test items per target. Test items did not appear in the familiarization. The full list of familiarization items appears in Appendix A.

[2] In most cases, randomization was implemented at the level of the participant by Paradigm software. However, the random order of the familiarization items in the randomized familiarization, ordered test condition was actually a pseudorandom order. Although this discrepancy was inadvertent, it was not expected to influence the results.

[3] In an effort to balance high and low experimental frequency targets, as well as high and low English frequency, across the eight target positions, some targets varied in terms of their experimental frequencies. Across all lists, the target /mp/ occurred slightly more often as low experimental frequency, and /st/ occurred more often as high experimental frequency.

[4] The term “target order” does not apply well to the randomized and ordered familiarization conditions. Compared to the other ordering conditions, the test item positions in the randomized and ordered familiarization conditions were not specific to a single consonant sequence. This can be seen in Table 1, where the first two items in both conditions are /paɪktəm/ containing /kt/ and /daʃpək/ containing /ʃp/. The ordered familiarization condition offered an alternative means of analysis: Rather than grouping test items by their order during test, they could be grouped by their order during familiarization. No statistical analysis changed as a result of this recoding, however.