The More the Merrier? The Impact of Talker Variability on Artificial Grammar Learning in Preschoolers and Adults

Federica Bulgarelli and Daniel J. Weiss

1. Introduction

A well-known challenge for hearing learners is that the input itself is typically produced by a variety of speakers with distinct vocal characteristics, reflecting variables such as gender, topic, and dialect. Therefore, the same speech category can differ greatly in its realization (e.g. Ladefoged, 1980; Peterson & Barney, 1952). Consequently, a prevalent assumption in much of the research on language acquisition has been that the task of the naïve learner is to acquire the linguistic system by overcoming the variability associated with individual speakers to ultimately abstract the core linguistic structure (e.g. Estes, 2012; Houston & Jusczyk, 2000).

Contending with speaker-variability can be cognitively demanding and lead to processing costs for learners across the lifespan. Specifically, changes in speaker from trial to trial have detrimental effects on word recognition (Mullennix et al., 1989; Creel & Bregman, 2011). For example, when 2-month-old infants are trained on words produced by multiple speakers, they have difficulty recognizing them after a brief delay (Jusczyk & Pisoni, 1992). When 3-to-5-year-old children are asked to match a familiar noun to a picture, word recognition is negatively impacted when the input is produced by multiple speakers. Similarly, when adults are asked to listen to and then type familiar words, receiving exposure to many speakers leads to lower accuracy for recognition (Mullennix et al., 1989). These processing costs extend to minimal talker variability, such as input produced by two speakers. When the input is produced by two speakers, learners may attempt to use them indexically and segregate the input based on speaker voice. For example, infants in a word segmentation task do not exhibit learning of the statistical regularities when they only hear two speakers during familiarization, even though both speakers produce the same speech stream (Estes & Lew-Williams, 2015).

While increasing talker variability may incur processing costs, it has also been found to help learners extract structural information in situations in which a single speaker is uninformative. Rost & McMurray (2009) discovered that increasing talker variability during familiarization using a switch task helped 14-month-old infants accurately learn novel minimal pairs that could not be acquired when the words were produced by a single talker. In the

* Federica Bulgarelli, Duke University, fedebul@gmail.com; Daniel J. Weiss, The Pennsylvania State University.

 Proceedings of the 43rd Boston University Conference on Language Development, ed. Megan M. Brown and Brady Dailey, 123-136. Somerville, MA: Cascadilla Press.

aforementioned speech segmentation task in which infants failed to segment a single stream of speech produced by two speakers, increasing the number of speakers in the input to eight led to successful segmentation (Estes & Lew-Williams, 2015). The benefits of talker-variability also extend to production. Four-year-olds were faster and more accurate in producing new vocabulary words when they were presented by different talkers during training (Richtsmeier, Gerken, Goffman, & Hogan, 2009). Adult L2 learners also benefit from increased talker variability during vocabulary learning, as evidenced by a positive correlation between their L2 production accuracy and the number of speakers heard during the training phase (Barcroft & Sommers, 2005). Contrasting the notion that talker variability is an impediment for learning, these studies suggest that talker variability may be helpful for abstracting relevant relationships that span talker-specific characteristics. We contend it is possible that the early difficulties imposed by talker variability may in fact result in more robust learning after acquisition.

The notion that overcoming difficulties can subsequently benefit learning is consistent with the concept of desirable difficulties (e.g. Bjork, 1994). A desirable difficulty is a manipulation of the learning environment (such as providing interference or varying learning conditions) that impairs immediate performance (Bjork, Dunlosky, & Kornell, 2011), but subsequently yields enhanced long-term retention and performance (e.g. Kornell & Bjork, 2008; Vlach, Sandhofer, & Kornell, 2008). Here we explicitly test whether talker variability during learning acts as a desirable difficulty for language learning. If so, we predicted talker variability should initially pose a challenge for learners, but also encourage long-term retention and generalization that are critical for language acquisition. We used an artificial grammar learning paradigm based on grammatical gender in which both adults and preschool-aged children were taught two categories of items, and then tested at three time points (spanning one week).

2. Experiment 1 – Artificial Language Learning in Adults

2.1. Participants

Participants were 145 monolingual English speakers (20 males, mean age = 18.85 years; SD = 1.49) recruited from an Intro to Psychology course who received course credit for participation. An additional 58 participants were recruited but excluded due to not returning for the Long Delay Test (n = 19), reporting English was not their first language (n=13), reporting near-native proficiency in a language other than English (n=1), not following instructions (n=6), experimenter error (n =7), or reporting low-effort (self-reporting 5 or below on a 10-point scale; n=12; see Mitchel & Weiss, 2010).

2.2. Stimuli

The stimuli consisted of 60 novel objects, 60 novel words (hereafter referred to as *nouns*). Twenty-four objects and word pairs were used during

familiarization, with the remainder were used at test. The objects consisted of black and white complex line drawings. Eight objects appeared in the stimuli used by Creel, Aslin, and Tanenhaus, (2008) and served as a template for creating the remaining objects for a previous experiment (Poepsel & Weiss, 2016). All of the novel nouns were created using English phonology, and ranged from 2 to 3 syllables, ending with a vowel. All final syllables were CV, while other syllables could be CV, CVC, or CCV. Half of the nouns ended in a front /a/ vowel (Category 1), while the other half ended with a back /ɑ/ vowel (Category 2). The articles were two monosyllabic artificial words, *sem* and *bol,* used in a previous study (Arnon & Ramscar, 2012). The auditory stimuli were recorded by 8 female native speakers of English who were trained to produce the vowel contrast. The articles and nouns were concatenated to form noun phrases (e.g., Sem barcha) with each article correlated with one of the two categories of nouns. This resulted in an underlying regularity between the article and vowel ending of the nouns, hereafter referred to as the category rule.

2.3. Procedure

Participants were seated in a sound attenuated booth and instructed that they would learn novel names for novel objects. During familiarization, participants viewed a picture of one of the objects on a computer screen, and heard a noun phrase describing the object. Each object was repeated 5 times during familiarization, in randomized order, for a total of 120 trials. Following familiarization, participants completed three test types (described below) that were conducted at three points in time. The first was presented immediately after familiarization (Immediate Test), and then after a 15-minute break (Short Delay Test) and finally a week after the original familiarization (Long Delay Test). Each of the three tests consisted of 12 test items, totaling 36 test items per test phase. For all of tests, participants viewed a single image on the screen, and heard two noun phrases, separated by a 1s pause. They were asked to choose the noun phrase that correctly described the image.

The Noun Test probed whether participants had learned the noun-object mappings. Participants viewed an object from familiarization and heard two noun phrases consisting of the correct article paired with either the correct or incorrect noun. The Article Test assessed learning of the article-noun pairings. Participants viewed an object and heard two noun phrases consisting of the corresponding noun paired with the correct and incorrect article. The Article Test used items that were heard during familiarization. The Generalization Test assessed participants' learning of the category rule and their ability to generalize to novel instances. During this test, participants viewed a novel object on the screen and heard two noun phrases. Both noun phrases contained a single novel label that could align with the category rule, paired with either article (e.g. *Sem peadla, Bol peadla,* see Figure 1). If participants inferred the underlying rule they would match the noun with the article that correctly matched the phonological category of the word. Twelve novel objects and their corresponding labels were used for each Generalization Test, resulting in 36

total novel objects. In addition to the experimental task, participants also completed an N-back task of working memory (adapted from Hakun & Johnson, 2017), and a Language History Questionnaire (Li, Sepanski, & Zhao, 2006).

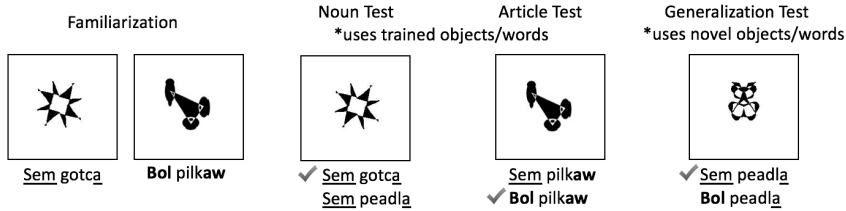

Figure 1. Examples of familiarization and test types for Experiment 1.

2.3.1. Conditions

Participants were assigned to one of three conditions: 1-Speaker, 2-Speaker or 8-Speaker. In the 1-Speaker condition, participants heard all stimuli presented by a single speaker, and hence no between-speaker variability. In the 2-Speaker condition, participants heard all of the familiarization phrases presented by two speakers. In the 8-Speaker condition, participants heard speech produced by all 8 speakers. At test, participants always heard one of the two speakers used in the 1-Speaker condition. For the 1-Speaker condition, the same speaker was used for familiarization and test.

2.4. Results and discussion

We used R (R Core Team, 2017) and the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar and R Core Team, 2017) to perform a repeated measures mixed effects analysis investigating performance across the Immediate, Short Delay and Long Delay Tests, based on condition. Three separate models were conducted, one for each type of test: Generalization, Noun, and Article. Test time, condition, the interaction between test time and condition, and N-back performance were fixed effects, and random effects for participant, with by-participant intercepts and by-participant random slopes for the effect of test time added to the model. Non-orthogonal contrast codes compared the 8-Speaker condition to the 1-Speaker condition and the 2-Speaker condition to the 1-Speaker condition. Orthogonal contrasts compared the Immediate to the Short Delay test, and the Immediate and Short Delay tests to the Long Delay test. Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity.

Results for all test types can be found in Figure 2 below. The repeated measures ANOVA for the Noun Test revealed an effect of Test Time such that performance at Time 1 and Time 2 was significantly higher than at Time 3, $t(284) = 7.34$, $p<.001$. There was no difference in performance between Time 1 and Time 2 ($t(284) = .17$, $p = .87$), suggesting that forgetting occurred only after the longer delay. Performance did not significantly differ across conditions (all

ps > .05), revealing no effect of talker variability on word learning. None of the interactions were significant (all *ps* > .05).

The repeated measures ANOVA for the Article Test revealed no effects of Test Time, such that performance did not change from the Immediate to the Short Delay test, $t(284) = .03$, $p=.97$; nor from the Immediate and Short Delay tests to the Long Delay, $t(284) = .67$, $p=.5$, suggesting that participants who performed above chance on the Immediate Test did not forget the article-noun mappings over the course of the experiment. Performance did not differ between the 1-Speaker and the 8-Speaker conditions, $t(141) = .52$, $p=.6$, but did differ between the 1-Speaker and 2-Speaker conditions, $t(141) = 2.82$, $p=.005$, such that performance in the 2-Speaker condition was significantly lower, $z = 2.82$, $p=.01$. This was true at all three time points: Immediate, $t(94) = 2.19$, $p=.031$; Short Delay, $t(94) = 2.13$, $p=.035$; and Long Delay, $t(94) = 2.08$, $p=.04$. None of the interactions were significant (all *ps* > .05).

The repeated measures ANOVA for the Generalization Test revealed no effect of Test Time, such that performance did not change from the Immediate to the Short Delay, $t(284) = 1.37$, $p=.17$, nor from the Immediate and the Short Delay to the Long Delay, $t(284) = 1.15$, $p=.25$. There was also no effect of Condition, such that performance did not differ between the 1-Speaker and the 8-Speaker condition, $t(141) = 1.24$, $p=.22$, nor between the 1-Speaker and the 2-Speaker condition, $t(141) = 1.24$, $p=.22$. There was, however, a significant interaction between the 1-Speaker and 8-Speaker condition, and the Immediate and Short Delay tests, $t(284) = 2.23$, $p=.026$. For the 1-Speaker condition, there was a trend towards performance decreasing from the Immediate (mean = 6.29, sd = 2.07) to the Short Delay test (mean = 5.54, sd = 1.81, $t(47) = 1.81$, $p=.077$). For the 8-Speaker condition, however, performance from the Immediate (mean = 6.02, sd = 2.36) to the Short Delay test (mean = 6.55, sd = 1.92) increased, though not significantly ($t(48) = 1.16$, $p=.25$). The Short Delay test was marginally above chance for the 8-speaker condition, $t(48) = 2.0$, $p=.051$.

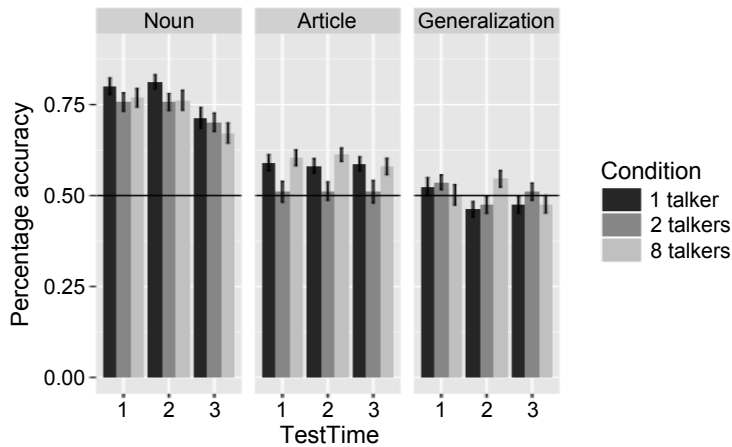

Figure 2. Adult performance across Test Type, Test Time and Condition.

2.4.1. Discussion

We investigated the role of talker variability on learning three aspects of an artificial grammar: nouns that described objects, article-noun pairings, and a category rule. Participants were exposed to 1-, 2-, or 8 speakers, and tested at three different time points. We predicted that if talker variability acted as a desirable difficulty for language learning, participants should exhibit a performance cost in the Immediate test for the 8-speaker relative to the 1-speaker condition, but over time learners in the 8-speaker condition should benefit with better retention. Further, if talker variability encourages learners to abstract the relevant information from the input, then only those in the 8-speaker condition should exhibit generalization to novel exemplars. The 2-speaker condition, however, may incur a more sustained cost as a result of limited talker variability, that might encourage learners to track speaker-specific utterances and not collapse information across speaker.

Contrary to these predictions, high talker variability condition did not negatively impact learning for the Noun and Article Tests, suggesting that multiple talkers did not incur a processing cost in this paradigm. Further, very little forgetting occurred overall, and retention of the learned information was not dependent on talker variability at training. In contrast to retention, generalization to novel exemplars was influenced by talker variability during training. As predicted, participants in the 1- and 2-speaker conditions did not exhibit generalization to novel exemplars at any of the test times. Participants in the 8-speaker condition, however, showed a trend towards successful generalization after the Short Delay, but performance was only marginally above chance. The vowel contrast used to identify category membership was rather subtle, perhaps resulting in the underlying category rule being too difficult or not sufficiently salient for learners. These results are consistent with the observation that learning grammatical gender can be rather challenging for second language learners (e.g. Dewaele & Véronique, 2001).

Likewise, presenting learners with low talker variability did not negatively impact learning of the nouns. However, it did negatively impact learning of the article-noun pairings as participants in the 2-Speaker condition did not exhibit learning in the Article Test at all time points. These results lend support to the idea that limited talker variability can interfere with learning.

3. Experiment 2 – Artificial Grammar Learning in Preschoolers

Given that three-to-four year old children are still actively engaged in language learning, Experiment 2 investigated the impact of talker variability on preschoolers. Previous studies have shown that both older (Braine et al., 1990; Brooks, Braine, Catalano, Brody, & Sudhalter, 1993) and younger (Gerken, Wilson, & Lewis, 2005) children are able to learn aspects of an artificial language containing grammatical gender. However, by age 3, children learning languages with grammatical gender show near mastery of this concept (e.g. Mariscal, 2009). This is the first study, to our knowledge, to investigate artificial

language learning with grammatical gender dependencies in this preschool age range.

3.1. Participants

Participants were 84 preschoolers (mean age = 43.18 months, SD = 3.52 months, 34 males) recruited from local daycares and preschools. Ten of the 84 participants were excluded from all analyses for perseverating on a single response during all tests. An additional 8 participants (beyond the 84) were recruited but excluded due to not meeting the criteria for age (n=4), not being willing to participate (n=2), or not understanding directions (n=2).

3.2. Stimuli

The stimuli consisted of 18 novel objects, 18 novel words (i.e., nouns, 6 used during familiarization, the rest used at test), 2 novel articles (as in Experiment 1) and 2 carrier phrases, which were sentence frames that preceded the novel words and articles (described below). The objects were taken from the NOUN database (Horst & Hout, 2015). All of the novel words were created using English phonology, and were all 2 syllables in length, ending with a vowel. In contrast to Experiment 1, half of the nouns ended in a back /ɑ/ vowel (Category 1), while the other half ended in a front /i/ vowel (Category 2). The carrier phrases used during the familiarization and test phases were "This is…" and "Is this…?". In addition to the experimental stimuli, a set of training stimuli was used to provide participants an example of how the task would unfold. The training objects included familiar items (dog, cat, house, bike), and the corresponding labels.

All auditory stimuli were recorded by the same 8 female native speakers of English used in Experiment 1. For half of the participants, Category 1 nouns were paired with *sem*, and for the other half they were paired with *bol.* The underlying regularity between the article and vowel ending of the nouns resulted in a category rule, as in Experiment 1.

3.3. Procedure

Participants were tested in a quiet area of their daycare or preschool. They were seated in front of a laptop computer and told that they would be playing a game during which they would be learning some new words corresponding to new objects. Prior to familiarization, participants saw two training trials. During the training trials, participants saw a picture of a familiar item (e.g. a dog), and heard a sentence correctly describing the picture (e.g. "This is a dog"). Following the training trials, the experimenter informed them that next they would see new pictures, and hear what they are called, and that they should listen and try to learn the new words.

During familiarization, participants viewed a picture of one of the novel objects on the screen, and heard a sentence describing the object. Each object

was repeated 4 times, for a total of 24 familiarization trials. Following familiarization, participants were instructed to answer questions by pressing a green button for 'yes' and a red button for 'no', and subsequently performed four practice trials with familiar objects. Next, participants completed three test types that were conducted at three time points, as in Experiment 1. The only difference relative to Experiment 1 was that the short delay was 10 minutes instead of 15 minutes.

Each testing phase consisted of 16 test trials. The three test types: Noun Test, Article Test and Generalization Test were the same as those used in Experiment 1, with a few modifications. Participants viewed an object on the screen and heard a single question about the object's name, labeling it either correctly or incorrectly (e.g. "Is this sem bemi?"). For each object, participants completed two test trials, one where the answer was 'yes' and one where the answer was 'no'. The Noun Test assessed whether participants had learned the noun-object mappings and consisted of four trials. The Article Test asked whether participants had learned the article-noun pairings, and also consisted of four trials. The Generalization Test assessed learning of the underlying rule by asking learners to generalize to novel instances. The test consisted of 8 trials using four novel objects to provide more chances for generalization. As in Experiment 1, participants were assigned to one of three conditions: 1-speaker, 2-speaker, or 8-speaker.

3.4. Results and discussion

The results for all three conditions at each time point can be found in Figure 3. Performance on the Noun Test did not exceed chance on any of the tests, regardless of condition (all *ps* >.05). Consequently, there were no effects of Test Time, $F(2,136) = 1.18$, $p=.31$, Condition, $F(2,70) = .03$, $p=.98$, Age, $F(1,70) = .02$, $p=.89$; and no interaction between Test Time and Condition $F(4,136) = .25$, $p=.91$.

The repeated measures ANOVA for the Article Test revealed no effect of Test Time, such that performance did not differ across test delays (all *ps* > .05). Performance between the 1-speaker and the 2-speaker condition was significantly different, $t(70) = 3.55$, $p<.001$, as participants in the 2-speaker condition outperformed those in the 1-speaker condition, $z = 3.55$, $p=.001$, while reaching above chance performance. Performance between the 1- and 8-speaker conditions was marginally different, $t(70) = 1.68$, $p=.097$, as learning trended higher in the 8-speaker condition, though never reached above chance. There was no effect of Age, $t(70) = .67$, $p=.5$. There was a marginal interaction between the 1- and 8-speaker conditions, and the Immediate and Short Delay tests, $t(139) = 1.73$, $p=.09$, with trends towards an increase in accuracy from the Immediate to the Short Delay test for the 1-speaker condition, but a decrease for the 8-speaker condition, though performance was never above chance. Further, the interaction between the 1- and 8-speaker conditions and the first two tests and the Long Delay test was also marginally significant, $t(139) = 1.85$, $p=.066$, with a trend towards improved performance for the 8-Speaker condition on the

Long Delay test, but not for the 1-Speaker condition. The other interactions were not significant (*ps* > .1).

Participants did not exceed chance on any of the tests of Generalization, regardless of condition (all *ps* > .05). The repeated measures ANOVA for the Generalization Test revealed no effects of test time (all ps > .05). Performance did differ between the 1- and 8-speaker conditions, $t(70) = 2.58$, p=.01, as performance was higher in the 8-speaker conditions, $z = 2.58$, p=.027, though none of the direct comparisons reached significance, all *ps*>.05. There was no difference between the 2- and 1-speaker conditions, no effect of age, nor were any of the interactions significant (all *ps*> .05).

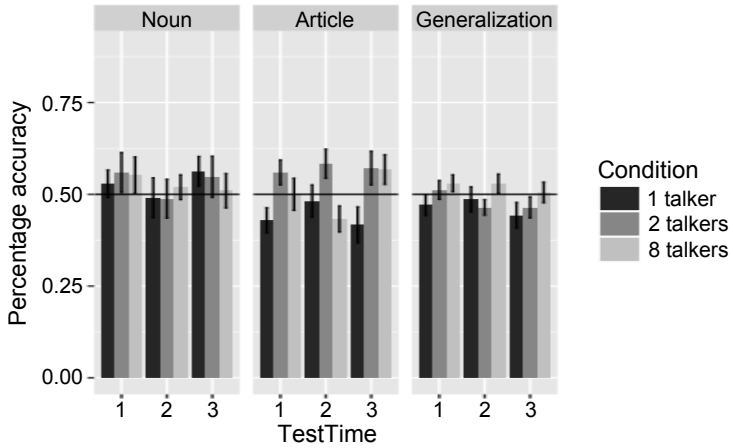

Figure 3. Preschooler performance across Test Time, Test Type and Condition

3.4.1. Discussion

In Experiment 2, we investigated whether talker variability acted as a desirable difficulty for language learning for preschool-aged children. The children were taught names for 6 novel objects divided into two categories. However, they exhibited difficulty learning the names, despite previous research reporting that 30 month-olds can fast map 6 novel names during a single experimental session (Golinkoff, Hirsh-pasek, Bailey, & Wenger, 1992). While previous research has reported that children can answer simple binary questions by 3 years of age (Ambridge & Rowland, 2013), those findings were in the context of native language queries (such as grammaticality judgments). In our study, answering yes-no questions may have been particularly difficult in the context of novel words.

Despite the lack of learning on the Noun Test, participants in the 2-Speaker conditions did exhibit some learning of the article-noun pairings. While it may seem counterintuitive that participants would learn the article-noun pairings but not object labels, this pattern of results is actually consistent with a recent study that found preliterate children were better able to learn article-noun pairs relative

to nouns in a similar grammar learning task focused on the effects of literacy (Havron, Inbal, Arnon, 2018). It may have also been easier for participants to judge the accuracy of the Article Test questions. The Noun Test required learners to adjudicate between two noun phrases heard during familiarization, whereas in the Article Test they could reject a noun and article unattested during familiarization. However, the novelty of the test items cannot entirely account for this pattern of results, as it was only evident in the 2-Speaker condition. While we predicted that receiving input from two speakers would result in interference for preschoolers, two speakers may have actually encouraged participants to learn two categories. Even though the speaker and category tokens were not perfectly correlated, hearing two speakers may have signaled to the learner the number of structures in the input (see General Discussion).

Participants did not exhibit generalization to novel exemplars regardless of the amount of talker variability they received during training. Considering the lack of learning evidenced in the Noun and Article Test, it is not surprising that participants did not generalize to novel exemplars.

4. General discussion

In two experiments, we investigated whether talker variability acts as a desirable difficult for language learning in adults and preschool aged children. Previous research on talker variability in language learning suggests that contending with talker variability can pose processing costs (Jusczyk & Pisoni, 1992; Mullennix et al., 1989). However, increasing talker variability can also help infants learn novel words (Rost & Mcmurray, 2009) that learners are unable to acquire in its absence. Our study aimed to investigate whether talker variability acts as a desirable difficulty for language learning. Participants were taught an artificial language comprised of novel names for novel objects, which were divided into two categories based on the word-final vowel for each word. Participants could therefore learn three aspects of the artificial language: the labels for each novel object, the noun-article pair, and an underlying rule linking the article with the word-final vowel. We found that adult learners in Experiment 1 exhibited significant learning of the object labels at all time points, regardless of the amount of talker variability presented during familiarization. Adults in the 1- and 8-speaker conditions also exhibited significant learning of the article-noun pairings, unlike participants in the 2-speaker condition. Only participants in the 8-speaker condition showed marginally above chance accuracy on the Generalization Test (on Short Delay test). In Experiment 2, preschool aged children exhibited difficulty learning the novel object names, never exhibiting above chance performance in any of the conditions. Contrasting adults in Experiment 1, only participants in the 2-Speaker condition exhibited learning of the article-noun pairings. Talker variability during familiarization did not impact generalization to novel exemplars, as performance on the Generalization Test also never exceeded chance.

The current study is the first, to our knowledge, to establish a pattern of retention for multiple components of an artificial language at multiple time points. Previous research has shown that adults can retain names of novel objects learned in a cross-situational statistical learning task (CSSL) after a week long delay (Vlach & Sandhofer, 2014). Here, participants acquired and retained both the novel object labels and the article-noun pairs, regardless of the amount of variability they received during training. These results add to a growing body of literature indicating that information learned during artificial language studies can lead to lasting memory traces (e.g. Arciuli & Simpson, 2012; Frank, Tenenbaum, & Gibson, 2013). This provides an important bridge between laboratory-based artificial language learning paradigms and challenges learners confront in the course of natural language acquisition.

In addition to testing whether high talker variability acted as a desirable difficulty, we also investigated the impact of minimal talker variability on artificial grammar learning. As previous research suggests that low talker variability interferes with segmenting a single stream of speech (Estes & Lew-Williams, 2015), we predicted that minimal talker variability would interfere with learning. In Experiment 1, adult learners seemed prone to interference when the input was presented by two speakers, as category identity was not perfectly correlated with talker. In Experiment 2 with preschoolers, however, the 2-speaker condition led to learning of the article-noun pairings, even though both speakers produced noun phrases from both categories (i.e, the speakers were not indexical cues). Preschoolers might use the amount of talker variability in the input to determine how many dimensions are relevant. According to this hypothesis, the presence of two speakers highlighted that two categories were present. However, when only one structure is present in the input (as in Estes & Lew-Williams, 2015), this can lead to interference. Thus, minimal talker variability may interact with both the number of structures in the input, and the age of the learner, as this reversal between adults and preschoolers was only evident for learning the article-noun pairs.

While our study did not provide further evidence that talker variability is indeed a desirable difficulty, it is possible this was due to methodological factors. For example, presenting input from only female speakers may not have introduced sufficient difficulty. Previous infant research suggests that the range of variability matters for generalization (e.g. Schmale, Cristia, & Seidl, 2012). Moreover, studies that found greater costs for higher talker variability in adult learners used 7 males and 8 females (Mullennix et al., 1989), providing even more variability in the input. Further, it is possible that the impact of talker variability is not uniform across tasks, time, and age. These factors are not orthogonal, such that the level of difficulty posed by the amount or range of talker variability may interact with the aspect of language to be learned. For example, talker variability has been found to positively impact vowel discrimination while negatively impacting word recognition in infants (Bergmann & Cristia, 2018; Bergmann & Tsuji, 2017). Therefore, future research might benefit from further exploration of the possibility that talker

variability acts as a desirable difficulty with an eye toward systematically investigating these possibilities.

The current experiments make a novel contribution to the literature by integrating research on talker variability with the notion of desirable difficulties in learning. While these experiments only provide weak support for the notion that talker variability may act as a desirable difficulty, the results suggest a series of future directions. An ongoing follow-up study is investigating whether using an artificial grammar with an easier vowel contrast with adults will facilitate more robust generalization, shedding light on the relationship between talker variability and the difficulty of the information to be learned. Further, in order to continue to investigate this question, future research needs to understand under which conditions talker variability leads to a processing cost, and the longevity of this processing cost. Once the factors that influence this processing cost are established, it may lead to clearer predictions for when a benefit for talker variability should be expected.

References

Ambridge, Benjamin, & Rowland, Caroline F. (2013). Experimental methods in studying child language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 149–168. http://doi.org/10.1002/wcs.1215

Arciuli, Joanne, & Simpson, Ian C. (2012). Statistical learning is lasting and consistent over time. *Neuroscience Letters*, *517*(2), 133–135. http://doi.org/10.1016/j.neulet.2012.04.045

Arnon, Inbal, & Ramscar, Michael. (2012) Granularity and the acquisition of grammartical gender: How order-of-acquisition affects what gets learned. *Cognition. 122*(3), 292-305.

Barcroft, Joe, & Sommers, Mitchell S. (2005). Effects of Acoustic Variability on Second Language Vocabulary Learning. *Studies in Second Language Acquisition*, *27*, 387–414. http://doi.org/10.1017/S0272263105050175

Bergmann, Christina, & Cristia, Alejandrina. (2018). Environmental Influences on Infants ' Native Vowel Discrimination : The Case of Talker Number in Daily Life, 1–18. http://doi.org/10.1111/infa.12232

Bergmann, Christina, & Tsuji, Sho. (2017). *Everyday input variability affects word recognition at 11 but not 16 months*. Retrieved from https://figshare.com/articles/Everyday_input_variability_affects_word_recognition_at_11_but_not_16_months/5492497

Bjork, Robert A. (1994). Memory and metamemory considerations in the training of human beings. *Metacognition: Knowing about Knowing.*

Bjork, Robert, A., Dunlosky, John, & Kornell, Nate. (2011). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, *64*(1), 120928131529005. http://doi.org/10.1146/annurev-psych-113011-143823

Braine, Martin, D. S., Brody, Ruth E., Brooks, Patricia J., Ross, Julie A., Catalano, Lisa, & Fisch, Shalom M. (1990). Exploring Language Acquisition in Children with a Miniature Artificial Language : Effects of Item and Pattern Frequency, Arbitrary Subclasses, and Correction. *Journal of Memory and Language*, *29*, 591-610.

Brooks, Patricia J., Braine, Martin D. S., Catalano, Lisa, Brody, Ruth E., & Sudhalter, Vicki. (1993). Acquisition of Gender-like Noun Subclasses in an Artificial

Language: The Contribution of Phonological Markers to Learning. *Journal of Memory and Language*, (32), 76–95.

Creel, Sarah C., Aslin, Richard N., & Tanenhaus, Michael K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664. http://doi.org/10.1016/j.cognition.2007.03.013

Dewaele, Jean-Marc, & Véronique, Daniel. (2001). Gender assignment and gender agreement in advanced French interlanguage: a cross-sectional study. *Bilingualism: Language and Cognition*, *4*(03), 275–297. http://doi.org/10.1017/S136672890100044X

Estes, Katherine G. (2012). Infants generalize representations of statistically segmented words. *Frontiers in Psychology*, *3*(OCT), 1–14. http://doi.org/10.3389/fpsyg.2012.00447

Estes, Katherine G., & Lew-Williams, Casey. (2015). Listening Through Voices: Infant Statistical Word Segmentation Across Multiple Speakers. *Developmental Psychology*, *51*(11), 1–12. http://doi.org/10.1037/a0039725

Frank, Michael C., Tenenbaum, Joshua B., & Gibson, Edward. (2013). Learning and Long-Term Retention of Large-Scale Artificial Languages. *PLoS ONE*, *8*(1), 1–6. http://doi.org/10.1371/journal.pone.0052500

Gerken, Louann, Wilson, Rrachel, & Lewis, Williams. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*(2), 249–268. http://doi.org/10.1017/S0305000904006786

Golinkoff, Roberta M., Hirsh-Pasek, Kathy, Bailey, Leslie M., & Wenger, Neill R. (1992). Children and Adults Use Lexical Principles to Learn New Nouns, *28*(1), 99–108.

Hakun, Johathan G., & Johnson, Nathan F. (2017). Dynamic range of frontoparietal functional modulation is associated with working memory capacity limitations in older adults. *Brain and Cognition*, *118*(August), 128–136. http://doi.org/10.1016/j.bandc.2017.08.007

Havron, Naomi, Raviv, Limor, & Arnon, Inbal. (2018). Literate and preliterate children show different patterns in an artificial language learning task. *Journal of Cultural Cognitive Science, 2*(1-2), 21-33.

Horst, Jessica, & Hout, Michael. (2015). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 1–17. http://doi.org/10.3758/s13428-015-0647-3

Houston, Derek M., & Jusczyk, Peter W. (2000). The Role of Talker-Specific Information in Word Segmentation by Infants, *26*(5), 1570–1582. http://doi.org/10.10371/0096-1523.26.5

Jusczyk, Peter W., & Pisoni, David B., Mullennix, John. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants, *43*, 253–291.

Kornell, Nate, & Bjork, Robert A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*(6), 585–592. http://doi.org/10.1111/j.1467-9280.2008.02127.x

Ladefoged, Peter. (1980). What Are Linguistic Sounds Made of? *Language*, *56*(3), 485–502.

Li, Ping, Sepanski, Sarah, & Zhao, Xiaowei. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–210. http://doi.org/10.3758/BF03192770

Mariscal, Sonia. (2009). Early acquisition of gender agreement in the Spanish noun phrase: starting small. *Journal of Child Language*, *36*(1), 143–171. http://doi.org/10.1017/S0305000908008908

Mitchel, Aaron D., & Weiss, Daniel J. (2010). What's in a face? Visual contributions to speech segmentation. *Language and Cognitive Processes*, *25*(4), 456–482. http://doi.org/10.1080/01690960903209888

Mullennix, John W., Pisoni, David B., & Martin, Christopher S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378. http://doi.org/10.1121/1.397688

Peterson, Gordon, & Barney, Harold. (1952). Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184. http://doi.org/10.1121/1.1906875

Pinheiro, Jose, Bates, Douglas, DebRoy, Saikat, Sarkar, Deepayan, & Team, R. C. (2017). Nlme: Linear and Nonlinear Mexied Effects models. R package version 3.1-128. 2016. *R software.*

Poepsel, Timothy J., & Weiss, Daniel J. (2016). The influence of bilingualism on statistical word learning. *Cognition*, *152*, 9–19. http://doi.org/10.1016/j.cognition.2016.03.001

Richtsmeier, Peter T., Gerken, Louann, Goffman, Lisa, & Hogan, Tiffany. (2009). Statistical frequency in perception affects children's lexical production. *Cognition*, *111*(3), 372–377. http://doi.org/10.1016/j.cognition.2009.02.009

Rost, Gwyneth C., & Mcmurray, Bob. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. http://doi.org/10.1111/j.1467-7687.2008.00786.x.Speaker

Schmale, Rachel, Cristia, Alejandrina, & Seidl, Amanda. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*, *15*(6), 732–738. http://doi.org/10.1111/j.1467-7687.2012.01175.x

Vlach, Haley A., & Sandhofer, Catherine M. (2014). Retrieval dynamics and retention in cross-situational statistical word learning. *Cognitive Science*, *38*(4), 757–774. http://doi.org/10.1111/cogs.12092

Vlach, Haley A., Sandhofer, Catherine M., & Kornell, Nate. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. http://doi.org/10.1016/j.cognition.2008.07.013

Proceedings of the 43rd annual Boston University Conference on Language Development

edited by Megan M. Brown
and Brady Dailey

Cascadilla Press Somerville, MA 2019

ISSN 1080-692X
ISBN 978-1-57473-096-8 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com