Talker variability shapes early word representations in English-learning 8-month-olds

Federica Bulgarelli[1] & Elika Bergelson[1]

[1] Duke University

Author Note

This is a preprint and has been accepted at Infancy as of December 2021. This work was supported by grants to EB (NIH-OD, DP5 OD019812-01) and FB (NIH-NICHD, F32 HD101216). We wish to thank all of the research assistants at Duke University who aided in recruitment and data collection, as well as those who helped record stimuli. The authors declare no conflicts of interest with regard to the funding source for this study.

Correspondence concerning this article should be addressed to Federica Bulgarelli, 417 Chapel Drive, Durham NC 27708. E-mail: fedebul@gmail.com

Abstract

Infants must form appropriately specific representations of how words sound, and what they mean. Previous research suggests that while 8-month-olds are learning words, they struggle with recognizing different-sounding instances of words (e.g. from new talkers), and with rejecting incorrect pronunciations. We asked how adding talker variability during learning may change infants' ability to learn and recognize words. Monolingual English-learning 7-to-9-month-olds heard a single novel word paired with an object in either a 'no variability,''within talker variability' or 'between talker variability' habituation. We then tested whether infants formed appropriately specific representations by changing the talker (Experiment 1a) or mispronouncing the word (Experiment 2), and by changing the trained word or object altogether (both experiments). Talker variability influenced learning. Infants trained with no talker variability learned the word-object link, but failed to recognize the word trained by a new talker, and were insensitive to the mispronunciation. Infants trained with talker variability dishabituated only to the new object, exhibiting difficulty forming the word-object link. Neither pattern is adult-like. Results are reported for both in-lab and Zoom participants. Implications for the role of talker variability in early word learning are discussed.

Keywords: talker variability, word learning, language development, infancy, switch task

Word count: 10770

Talker variability shapes early word representations in English-learning 8-month-olds

Introduction

Words sound slightly different each time they are said, due to factors such as gender, age, topic, register and dialect (Liberman, Coopers, Shankweiler, & Studdert-Kennedy, 1967). As a result, word learning requires forming appropriately specific representations of how words sound, as well as what they mean. In some ways, infants rapidly rise to this challenge, understanding common nouns (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012) and showing language-specific phonetic knowledge (Polka & Werker, 1994; Werker & Tees, 1984) before age one. In other ways, young infants struggle with word-form recognition. Specifically, young infants have difficulty recognizing new instances of spoken words, (e.g. produced by a novel talker (e.g. Houston & Jusczyk, 2000), in a new affect (e.g. Singh, Morgan, & White, 2004), or in a different accent (e.g. Schmale & Seidl, 2009)). They also have difficulty correctly rejecting incorrect instances of spoken words (e.g. when they are mispronounced (e.g. Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015; Singh, 2008)). Here we ask whether and how hearing different-sounding examples of a word during training shapes what infants attend to in the earliest phases of learning a new word.

Word-form recognition

As noted above, a critical component of word learning is being able to recognize novel instances of a word. Around 7 months of age, infants have trouble with this, suggesting relatively fragile representations of learned words (see Singh, 2008). A well-established line of research has investigated early word form recognition by playing infants lists of common words (e.g. bike, tree, pear) in the absence of their visual referents, and subsequently asking whether infants recognize those words when the surface form changes, e.g. when the word sounds different because it is produced by a new talker or in a new affect.

Tested with this approach, 7.5-month-olds fail to recognize words they initially heard by a male talker when they're spoken by a female talker (Houston & Jusczyk, 2000), or words initially heard in a single affect when spoken in a new affect (Singh, Morgan, & White, 2004). With a few more months' learning and experience, infants overcome these overly constrained representations of what words should sound like, as by 10.5 months, they succeed at recognizing trained words when spoken by a new talker or in a new affect (Houston & Jusczyk, 2000; Singh, Morgan, & White, 2004). While some research suggests that contending with input from multiple talkers makes word-form recognition *more* difficult across the lifespan (Jusczyk, Pisoni, & Mullennix, 1992; Mullennix, Pisoni, & Martin, 1989; Ryalls & Pisoni, 1997), more variable training has also been found to improve infants' abilities to recognize words that differ in their surface form. For instance, hearing words from multiple talkers or in multiple affects in a training phase has been shown to help 7.5-month-olds recognize those words when they hear them from a new talker or in a new affect (Houston, 1999; Singh, 2008).

Another facet of word-form recognition is learning when the sounds of the word have changed *enough* to possibly signal a change in meaning. Around 5 months of age, infants actually fail to detect mispronunciations of their own name (Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015). By 11 months of age, however, infants prefer correct over mispronounced versions of common nouns (Swingley, 2005). In fact, at 7.5 months of age (i.e. in between these ages), infants accept 'gare' as an instance of 'pear' in the absence of acoustic variability (Singh, 2008). Here too, variability during training helps 7.5-month-old infants reject single-phoneme mispronunciations; that is, hearing 'pear' with high affect variability during training leads infants to reject 'gare' as an instance of 'pear' (Singh, 2008).

Taken together, 7.5-month-olds' representations of words are sometimes overly specific, leading them to not recognize words that differ in their surface forms (e.g. new talker), and sometimes overly broad, leading them to accept incorrect pronunciations. In both cases, acoustic variability has been shown to help infants focus on which aspects of the acoustic signal are important to attend to in order to recognize familiar words (e.g. Singh, 2008).

Word learning

Beyond helping infants recognize viable instances and reject incorrect instances of familiar words, a separate line of research has also found that increasing talker variability can help older infants learn new words in the lab (Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009; see also Richtsmeier, Gerken, Goffman, & Hogan, 2009 for a similar effect in preschoolers). Lab studies find that 14-month-olds exhibit difficulty learning two new similar-sounding words for new objects (Stager & Werker, 1997) in the absence of talker variability. A paradigm commonly used to study this is the Switch task, in which participants are familiarized to two word-object pairs (object-a and word-a; object-b and word-b) until habituation, and tested with a 'switch' of this pairing (e.g. object-a with word-b). An increase in looking time to the 'switch' trial is taken to indicate learning of the word-object association (Werker, Cohen, Lloyd, & Stager, 1998). When the novel words sound sufficiently distinct (e.g. lif and neem), 14-month-olds increase their looking time, noticing the switch (Werker, Cohen, Lloyd, & Stager, 1998). However, when these words are minimal pairs (i.e. they differ by one speech-sound, e.g. bih and dih), infants fail to notice the switch (Stager & Werker, 1997). Critically, this failure is not due to 14-month-olds' inability to *hear* the difference between the words' sounds, but rather their inability to link similar-sounding words to distinct objects (Stager & Werker, 1997).

Follow-up work has found a variety of manipulations that help 14-month-olds succeed in the (more challenging) minimal pair switch task (e.g. Fennell, 2012; Fennell & Waxman, 2010; Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009). Most germane here, McMurray and colleagues (Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009, 2010 Exp 3) proposed that, as above for word-form recognition, increasing talker variability may draw learners' attention to the features of words that remain consistent. That is, 'task-irrelevant' variability may highlight *relevant* differences between words, i.e. the difference in their speech sounds (cf. Gogate & Hollich, 2010 on 'invariance detection'; Apfelbaum & McMurray, 2011).

Supporting this idea, Rost and McMurray (2009) first replicated Stager and Werker (1997) using a single token from a single talker, finding too that 14-month-olds fail to learn the word-object links. However, they then showed that training with between-talker variability (i.e. 18 different talkers, half male and half female) led infants to notice the word-object switch (Rost & McMurray, 2009; see Hohle, Fritzsche, Meb, Philipp, & Gafos, 2020 for a replication in German). Similarly, training with within-talker variability (i.e. a single highly-variable talker) also led 14-month-olds to succeed (Galle, Apfelbaum, & McMurray, 2015). Notably, manipulating a phonemically contrastive dimension (e.g. voice-onset time) did not lead 14-month-olds to notice the switch (Rost & McMurray, 2010), nor did training one word with a set of female talkers and the other with a set of male talkers, presumably because this doesn't highlight how the words differ for the same set of talkers (Quam, Knight, & Gerken, 2017). Taken together, previous lab studies suggest between- and within- talker acoustic variability can help 14-month-olds learn novel minimal pairs, by encouraging them to attend to relevant features of those words.

Current studies

Taken together, acoustic variability helps infants realize which aspects of the acoustic signal are important to attend to, both for appropriately recognizing instances of familiar words around 7-8 months of age (e.g. Singh, 2008), and for learning novel minimal pairs around 14 months (Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009; see also Quam & Creel, 2021 for a review). But what role does talker variability play for younger infants during the initial process of learning novel words? At 8 months of age, infants are learning words and forming relatively robust word-object links. For example, Bergelson and Swingley (2012) showed that infants between 6 and 9 months of age look at images of foods and body parts when hearing them labeled aloud. Similarly, Tincoff and colleagues showed that 6 month old infants can link words to their specific one-to-one associations (e.g. Mommy referring only to the infant's mother and not to other female adults, Tincoff & Jusczyk, 1999), and to categories of objects (e.g. foot referring to other people's feet, Tincoff & Jusczyk, 2012). However, at this age, infants still exhibit difficulty (1) generalizing to surface-level (i.e. non-phonemic) changes (Houston & Jusczyk, 2000; Singh, 2008; though not always, see Bergelson & Aslin, 2017), and (2) rejecting phonemic changes (e.g. mispronunciations). In what follows, we extend previous research testing familiar word recognition (e.g. Singh, 2008) and ask how talker variability shapes the information that younger infants (8-month-olds) attend to in the process of forming new word-object links.

Since the two-word switch task is not generally used before 14 months, we used the simplified one-word version previously used with 8-month-olds (Stager & Werker, 1997; Werker, Cohen, Lloyd, & Stager, 1998) in which infants are habituated to a single novel word-object pairing (e.g. "lif" or "neem"). In this simplified 1-object switch task, 8-month olds dishabituated when the trained object was paired with a novel word, or when the trained word was paired with a novel object (Werker, Cohen, Lloyd, & Stager, 1998). The fundamental assumption of this method is that infants look longer when a critical component of the word-object link they've been habituated to has been altered, relative to their looking when presented with the same word-object link from the habituation phase. Of course, word learning is a complex process that typically unfolds over thousands of experiences with utterances and interactions in the world. Here we isolate an extremely limited version of this learning process. This approach relies on infants' nascent knowledge of their native language phonology, alongside their visual and auditory discrimination and categorization skills.

In the current study, infants were taught a new word-object pair in one of three *habituation conditions*: no talker variability, within-talker variability or between-talker variability. Notably, the within- and between- talker variability used here is similar to what infants are exposed to in their daily lives (Bulgarelli, Mielke, & Bergelson, in press). Once habituated, infants were then tested to see if they noticed three types of changes relative to a *same trial*: a *critical test trial* and two *control trials*. In Experiment 1, this critical trial tests whether infants noticed when they heard a brand new talker (of another gender) produce the trained word. A new talker is a non-criterial change to the word-object link; hearing the word from a new talker should not be noteworthy. In Experiment 2, the critical trial probed whether infants noticed when the word was mispronounced (i.e. the vowel in the word changed). In contrast to a new talker, a change in a single phoneme of a word is criterial, as the altered word could refer to a new object (e.g. ball vs. bell). In both experiments, this critical test trial was followed by two control trials probing whether infants noticed when they were presented with a brand new word or a brand new object (each from a familiar talker) instead of the trained word and object. These control trials were intended to be confirmatory: they are both large changes that inarguably break the word-object link.

Given that infants at this age fail to recognize familiar words produced by new talkers when trained without talker variability (Houston & Jusczyk, 2000), we may find that regardless of habituation condition, they consider a talker change (Experiment 1) to be a notable divergence from the trained word-object link, leading them to dishabituate. In contrast, introducing talker variability (within- or between- talkers) in the habituation phase may highlight the irrelevance of talker for word identity. In this case, infants in the within- or between- talker-variability habituation conditions would show no change in their behavior when the talker switches at test. Similarly, given that infants at this age (incorrectly) accept mispronunciations of familiar words (Singh, 2008), we may find that regardless of habituation condition, they do not consider a mispronunciation noteworthy, i.e., fail to dishabituate (Experiment 2). In contrast, if talker variability during habituation highlights the importance of phonemic constancy for word identity (Rost & McMurray, 2009), then infants in the within- or between- talker-variability habituation conditions may instead dishabituate to the mispronunciation at test. Based on previous research, we predict that in the control trials, infants in all three conditions across both Experiments will notice (i.e. dishabituate) when the word or the object change. The results of this study carry implications regarding features of infants' input that may – naturally or through intervention – serve to shape early word learning.

Experiment 1a

In Experiment 1a, we test whether talker variability during habituation to a novel object-word pair influences looking times when infants are presented with an instance of the trained word produced by a new talker. By hypothesis, infants who have formed a properly-scoped link between the word and object should find a talker change unremarkable, because a change in talker does not break the word-object link.

Methods

The preregistration (https://osf.io/acrsp), as well as all stimuli, data, and code used to create this manuscript are posted through the Open Science Foundation (OSF): https://osf.io/xwsnm/. A power analysis prior to data collect (see preregistration) found that for a within- and between-subject analysis, a sample of 18 participants per condition would be sufficient to achieve .95 power to detect a medium effect size (.25). This sample size is consistent with previous studies using the Switch paradigm, which result in a moderate effect size (Cohen's d = .32, based on Tsui, Byers-Heinlein, & Fennell, 2019), and is what we use here.

Participants. Our final sample was made up of 54 7- to 9-month-old infants (26 female, 28 male, M_{age} = 7.98 months). All participants were full term (40 +/- 3 weeks), monolingual (parents did not report >25% exposure to a language other than English), and had no history of hearing or vision problems. Participants were recruited from the broader area surrounding a university in the Southeastern United States. Parents provided consent on behalf of themselves and their infants, and were compensated for travel ($5 or $10 depending on distance traveled) and participation (a child-focused thank you gift, e.g. a book, small toy, or t-shirt). 76% of the infants were White or Caucasian, 4% were Black or African American, and 20% identified as other or multiracial. Maternal education ranged from some high school to advanced degree (some high school: n=1; high school degree: n = 1; some trade school, professional training, or college: n = 2; vocational, trade, or technical diploma: n = 1; associate or bachelor's degree: n = 24; advanced degree: n = 24). An additional 15 infants were excluded due to fussiness (N = 6), technical difficulties (N = 3), parental interference (N = 2), not meeting our language exposure criteria (N = 2), or prematurity (N = 2). The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board at Duke University.

Design. The experiment consisted of a single-word switch task, wherein participants were habituated to a single word-object pair in one of three conditions: *No-Talker-Variability*, *Within-Talker-Variability*, or *Between-Talker-Variability*[1]. In the *No-Talker-Variability* condition, infants heard a single prototypical child-directed token of the novel word produced by a single female talker. In the *Within-Talker-Variability* condition, infants heard 12 highly-variable tokens produced by a single female talker. Finally, in the *Between-Talker-Variability* condition, infants heard 10 different female talkers produce the novel word. The test phase queried what changes to the word-object link infants noticed. All infants saw four types of test trials: a *Same* trial and three Switch trials: a *Talker Switch* trial, a *Word Switch* and a *Picture Switch*; see Figure 1.

Stimuli. Stimuli consisted of four familiar warm-up items (apple, ball, shoe, dog), and two novel items (object1 - a kitchen tool, object2 - a dog toy) and their corresponding

[1] The *No-Talker-Variability* condition was run in full first, in order to establish that our instantiation of the single-item switch task worked in a condition where we had a strong prediction for the *Talker Switch* test trial (see preregistration). Thereafter, the *Within-Talker-Variability* and *Between-Talker-Variability* conditions were run in parallel, with random assignment of infants to condition.

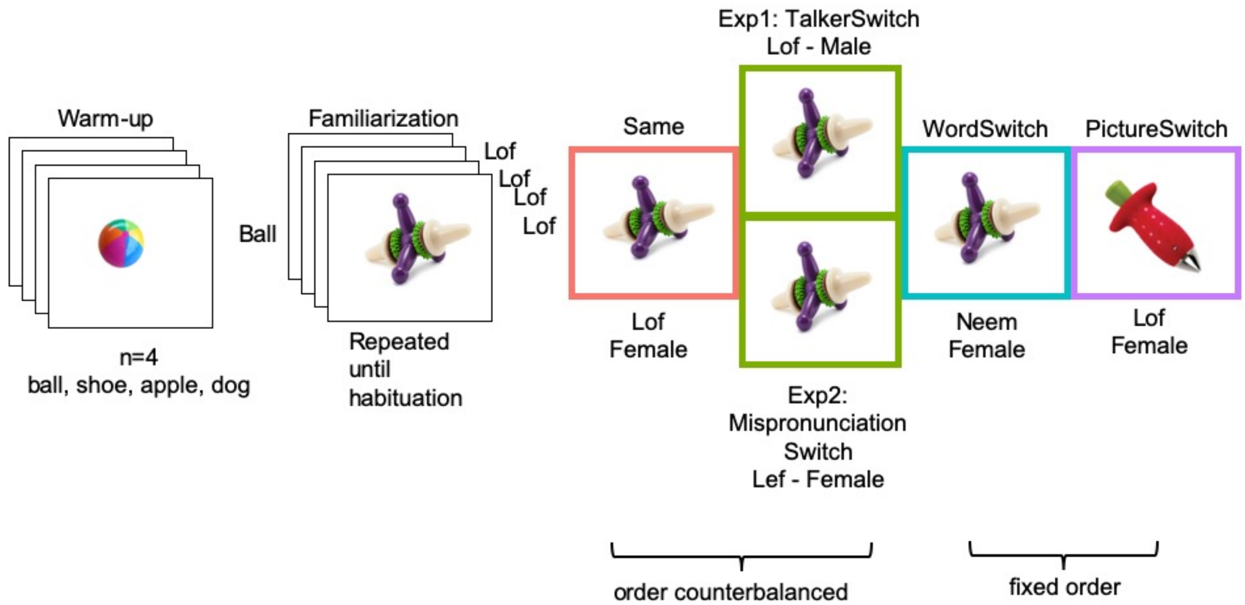

Figure 1. Experimental procedure. Colored boxes correspond to data in subsequent figures

labels (‘neem’ and ‘lof’); as well as an animated attention-getter paired with a jingle.

Visual stimuli consisted of animated videos of the warm-up items and novel objects. The videos showed the objects looming on the screen, ranging from 50-90% in height and 30-50% in width of the display.

Auditory stimuli consisted of recordings of the warm-up items and novel words for the habituation and test phase. Each word was recorded by 10 female young adults (used in habituation and test) and 2 male young adults (used only at test). Our auditory stimuli deliberately maximized acoustic differences stemming from within- and between- talker variability, our main variable of interest. To achieve this, each talker recorded each novel word six times and each familiar word three times in child directed speech, and recorded each novel word nine additional times by systematically varying the overall pitch (normal/high/low), pitch contour (rising/flat/falling), and duration (normal/short/long) of the word (cf. Galle, Apfelbaum, and McMurray (2015)); two female talkers did the same for the warm-up items. By recording stimuli in this way, we introduced naturalistic talker variability, which varied in multiple dimensions by design. Each token was then spliced and embedded in silence, resulting in 2s long sound files. These sound files were then normalized to a mean intensity of 71 dB, see Supplementals for additional details and the OSF link above to hear and see all stimuli.

Caregiver questionnaires. Caregivers filled out three questionnaires: (1) the MacArthur-Bates Communicative Development Inventory (CDI), Words and Gestures Form (Fenson et al., 1994), a vocabulary checklist where parents indicate words their child understands or says; (2) a language exposure survey asking about the varieties of English and any other languages participants may be exposed to; and (3) a demographics questionnaire including information such as age and gender. See Supplementals for results from the CDI and language exposure survey.

Procedure. After consent and questionnaires, infants and caregivers were escorted to the testing room, where participants sat in their caregiver's lap facing a 43" monitor within a 7.5 x 8ft sound-attenuated booth. Caregivers listened to music over headphones, ensuring they would not hear the experimental stimuli and influence their children's behavior. An experimenter sat outside the booth and live-coded infants' looks to the monitor via button press. Critically, the experimenter had access to the child's looking behavior, but could not hear or see the stimuli inside the booth.

The experiment was run using Habit 2 (Oakes, Sperka, DeBolt, & Cantrell, 2019). Each trial began with an attention-getter directing infants' gaze to the monitor. All trials lasted up to 14 seconds (i.e. 7 instances of the word-object pair), and remained on the screen as long as participants were looking at them, or until the maximum time had elapsed. If participants looked away for more than 2s after looking at the screen for at least 1s, the trial advanced.

Warm-up trials. The experiment began with four warm-up trials to introduce infants to the idea that this task concerned objects and their labels, as the use of referential cues has been shown to help older infants succeed in the challenging minimal pair task

(Fennell & Waxman, 2010). In these trials, participants saw a looming familiar object while hearing it labeled aloud, see Figure 1.

Habituation phase. In the habituation phase, participants viewed a video of a novel object looming on the screen while hearing the corresponding novel word. The habituation phase continued until participants reached our habituation criteria: when looking time to the last four trials was half as long as looking time to the first four trials, using a sliding window (Casasola & Cohen, 2000); and could last between 5 and 30 trials. All participants met our habituation criteria.

Test trials. After the Habituation phase, infants were advanced to the test phase, which consisted of four trials: a *Same* trial and three *Switch trials*, each lasting up to 14s. The *Same* trial repeated a token used during habituation. The *Talker Switch* trial was the critical test trial. This trial repeated a single token of the correct word by a previously-unheard male talker. This tests infants' ability to recognize the recently learned word with a talker (and gender) change, which does not violate the word-object link. The other two switch trials were control trials. In the *Word Switch* trial, infants saw the trained object and heard a brand new word (e.g. 'lof' if they were trained on 'neem'). For the *Picture Switch* trial, infants saw a brand new object while hearing the trained word (e.g. if they were trained with object-1 as 'neem' they saw object-2 and heard 'neem'). These control trials query whether infants detect the violation of the word-object link. The *Same* and *Talker Switch* trials occurred first, and were counterbalanced across participants; these were followed by the *Word Switch* and the *Picture Switch* trial in a fixed order, see Figure 1.

Counterbalancing. One of 2 female talkers were used for familiarization in the *No-Talker-Variability* condition and the *Within-Talker* variability condition, and for test trials across all three conditions. The specific talker was counterbalanced across participants. Ten female talkers (including the 2 just mentioned) were used for familiarization in the Between Talker Variability condition. To facilitate counterbalancing across participants, word-object pair and talker were yoked. For example, all participants who learned word-object pair 1 (e.g. neem and the kitchen tool) always heard female-1 for the *Same*, *Word Switch* and *Picture Switch* test trials, and male-1 for the *Talker Switch* trial, regardless of talker variability condition during habituation. For the preceding warm-up trials and habituation phase, those in the *No-Talker-Variability* and *Within-Talker-Variability* conditions therefore also heard female-1, while those in the *Between-Talker-Variability* condition heard female-1 in addition to other female talkers.

Results

Analysis Plan. We used RStudio (RStudio Team, 2019) and R [Version 4.0.2; R Core Team (2017)] to generate this manuscript, along with all figures and analyses. See Supplementals for specific library details; all libraries are cited in the references.

For our main analysis, we conducted mixed effects regressions using lme4 (Bates, Mächler, Bolker, & Walker, 2015) to test whether looking time to the *Switch* test trials (Talker Switch, Word Switch and Picture Switch) differed from the Same test trial, by habituation condition. We included effects for trial type, condition (*No-Talker-Variability*, *Within-Talker-Variability*, *Between-Talker-Variability*) and the interaction between them. To account for possible stimuli or order idiosyncrasies, we included random intercepts for word-object pair (which also includes talker, by design) and trial-order. We further included by-Subject random intercepts in the model. Thus, the model formula was as follows:

LookingTime ~ TestTrialType * HabituationCondition + (1|Subj) + (1|Word-object-pair) + (1|TestTrialOrder)

Since the Same test trial served as our baseline, our trial type contrasts were set up to compare looking time between the Same test trial and each of the three Switch trials (Talker Switch, Word Switch, Picture Switch) separately. To test the effects of talker variability during training, we used orthogonal contrast codes for the three habituation conditions (no-, between-, and within-talker variability). Given that previous research has found within- and between- talker variability have similar effects on word learning (Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009; Tsui, Byers-Heinlein, & Fennell, 2019), one of our sets of contrasts combines them, i.e. compares the *No-Talker-Variability* condition to the two conditions featuring talker variability together. The other set of contrasts compares the *Between-Talker-Variability* and *Within-Talker-Variability* conditions to each other. Given the nature of our analysis, we do not report omnibus effects for each variable, and instead report results for our specific contrasts of interest. Thus, based on how the contrasts were set up, an interaction between our trial type contrasts and the habituation condition contrasts would indicate that differences in looking time between specific trials (e.g. Same vs. Talker Switch) differ by habituation condition.

Habituation Results. Before conducting our main analysis of the test trials, we first analyzed whether habituation times differed by habituation conditions. Across all three habituation conditions, infants habituated after an average of 12.96(SD = 4.71) trials. However, this differed by habituation condition, $F(2, 51) = 4.31$, $MSE = 19.72$, $p = .019$; participants in the *No-Talker-Variability* condition habituated in 14.61(SD = 4.51) trials, which did not differ significantly from those in the *Between-Talker-Variability* condition (mean = 13.78, SD = 5.22, $\Delta M = 0.83$, 95% CI $[-2.47, 4.14]$, $t(33.31) = 0.51$, $p = .612$), but did differ significantly from participants in the *Within-Talker-Variability* condition who exhibited significantly faster habituation (mean = 10.50, SD =3.40, $\Delta M = 4.11$, 95% CI $[1.40, 6.83]$, $t(31.59) = 3.09$, $p = .004$). Participants in the *Within-Talker-Variability* condition also habituated faster than those in the *Between-Talker-Variability* condition, $\Delta M = 3.28$, 95% CI $[0.28, 6.28]$, $t(29.23) = 2.23$, $p = .033$.

Test Trial Results. Results for the test trials are visualized in Figure 2 (1a panels); full model results, including Cohen's d can be found in Table 1, t and p values are also reported in text. We report main effects for each contrast first, followed by the interactions.

There was no main effect of habituation condition: looking time did not differ *overall* between the *No-Talker-Variability* condition and the two conditions featuring talker variability (t = 1.15, p = .258), nor between the *Within-Talker-Variability* and the *Between-Talker-Variability* conditions (t = 1.17, p =.248).

There was a significant main effect of trial, such that infants across all conditions increased their looking time to the control switch trials, i.e. the *Word Switch* trial ($M_{\text{WordSwitch}}$ = 7.14s, SD = 4.60), and the *Picture Switch* trial ($M_{\text{PictureSwitch}}$ = 8.68s, SD = 3.91) relative to the *Same* trial (M_{Same} = 5.55s, SD = 3.45; Same vs. Word Switch: t = 2.70, p = .008, Same vs. Picture Switch: t = 5.32, p < .001). However, looking time to the critical *Talker Switch* test trial did not differ from looking time to the *Same* trial ($M_{\text{TalkerSwitch}}$ = 6.06s, SD = 3.73), t = 0.87, p = .386.

No significant interactions included the contrast comparing the *Within-Talker-Variability* vs. *Between-Talker-Variability* conditions (all ps > 0.18). This suggests that performance on this task was not predicted by the type of talker variability that infants received during habituation in those conditions, i.e. between vs. within talkers. Given this, in what follows we do not report means for the between- and within- talker variability condition separately in text, though they can be found in Figure 2 (1a panels) and in footnotes.

There were significant interactions between looking time to different trial types for participants in the *No-Talker-Variability* condition vs. the two conditions featuring talker variability together. Specifically, looking time to the *Talker Switch* trial vs. *Same* trial differed depending on whether the condition featured talker variability (t = 3.36, p = .001): *Talker Switch* trial looking-time was significantly higher than *Same* trial looking-time in the *No-Talker-Variability* condition (M_{Same} = 4.94s, SD = 2.73, $M_{\text{TalkerSwitch}}$ = 8.26s, SD = 3.45), but did not significantly differ for the talker variability conditions together[2] (M_{Same} = 5.85s, SD = 3.76, $M_{\text{TalkerSwitch}}$ = 4.97s, SD = 3.40); $t(35)$ = 1.33, p = .193. This suggests that only

[2] Within-talker variability: M_{Same} = 4.89, $M_{\text{TalkerSwitch}}$ = 4.97. Between-talker variability: M_{Same} = 6.81, $M_{\text{TalkerSwitch}}$ = 4.96.

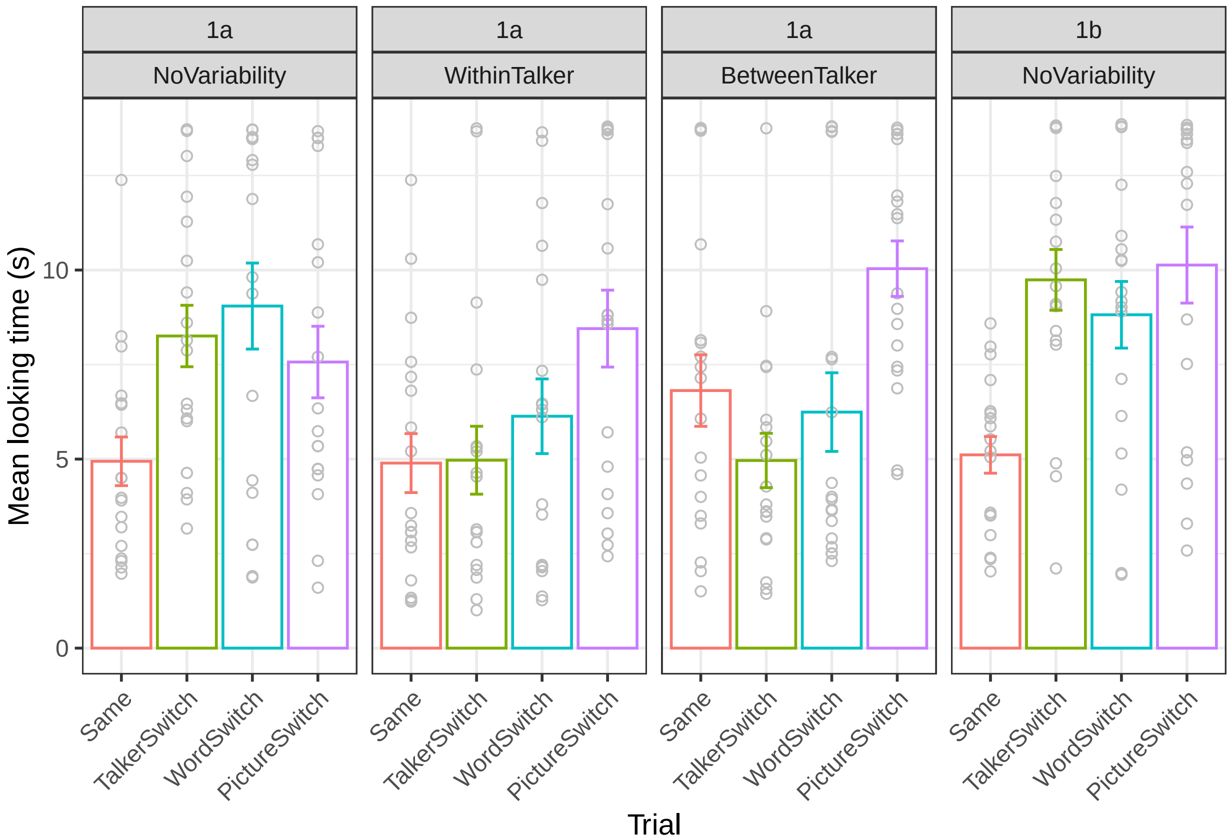

Figure 2. Results for Experiment 1a and 1b. Bars depict mean looking time (y-axis) across test trials for participants in all three conditions in 1a and for the No Variability condition in 1b (x-axis). Circles indicate individual data points, error bars reflect standard error. Participants in the No-Talker-Variability condition (in both 1a and 1b) dishabituated to all three Switch trials (Talker, Word, and Picture). Participants in both Talker Variability conditions only dishabituated to the Picture Switch trial.

Table 1

Fixed effects and Cohen's d for Experiment 1a model. '/' in predictor name indicates the specified contrast (e.g. Same/TalkerSwitch compares looking time to Same vs TalkerSwitch trial); ':' indicates an interaction between specified contrasts. SE is pooled for each predictor

term	estimate	std.error	statistic	p.value	d
(Intercept)	6,890.41	1,071.50	6.43	0.059	NA
Same/TalkerSwitch	513.22	589.81	0.87	0.386	0.141
Same/WordSwitch	1,591.52	589.81	2.70	0.008	0.436
Same/PictureSwitch	3,135.13	589.81	5.32	<.001	0.859
NoVariability/TalkerVariability	843.43	736.44	1.15	0.258	0.327
WithinTalker/BetweenTalker	497.21	425.44	1.17	0.248	0.334
Same/TalkerSwitch:NoVariability/TalkerVariability	4,200.75	1,251.17	3.36	<.001	0.543
Same/WordSwitch:NoVariability/TalkerVariability	3,771.47	1,251.17	3.01	0.003	0.487
Same/PictureSwitch:NoVariability/TalkerVariability	-764.53	1,251.17	-0.61	0.542	-0.099
Same/TalkerSwitch:WithinTalker/BetweenTalker	-962.69	722.36	-1.33	0.185	-0.215
Same/WordSwitch:WithinTalker/BetweenTalker	-904.36	722.36	-1.25	0.212	-0.202
Same/PictureSwitch:WithinTalker/BetweenTalker	-166.53	722.36	-0.23	0.818	-0.037

after training a word-object link with talker variability do infants treat a talker change as unremarkable (i.e. they did not dishabituate to it, relative to the originally presented word-object link in the *Same* trial).

Looking time to the *Word Switch* control trial vs. *Same* trial also differed across conditions that featured talker variability, $t = 3.01$, $p = .003$: looking time to the *Word Switch* trial was significantly higher in the *No-Talker-Variability* condition ($M_{\text{Same}} = 4.94$s, $SD = 4.94$; $M_{\text{WordSwitch}} = 9.05$s, $SD = 4.82$; $t(17) = -3.59$, $p = .002$), but did not significantly differ in the within and between talker variability conditions together[3], (M_{Same} $= 5.85$s, SD $= 3.76$; $M_{WordSwitch} = 6.19$s, 4.24; $t(35) = -0.45$, $p = .656$). This suggests that training with talker variability led infants to (erroneously) ignore a change in object-label. That is, infants' looking to the screen did not increase significantly when the objects' label changed to a brand new word in the two conditions featuring talker variability, but did increase after word-object training without talker variability (i.e. in the *No-Talker-Variability* habituation condition).

Lastly, looking time to the *Picture Switch* control trial vs. the *Same* trial did not differ across conditions, regardless of whether they featured talker variability, t = -0.61, p = .542. That is, looking time to the *Picture Switch* control trial was significantly higher than that for the *Same* trial for participants in all conditions (*No-Talker-Variability*: $M_{Same} = 4.94$s, SD $=$ 2.73; $M_{PictureSwitch} = 7.57$s, SD $= 4.01$; $t(17) = -3.37$, $p = .004$; *Talker-Variability*: $M_{Same} =$ 5.85s, SD $= 3.76$; $M_{PictureSwitch} = 9.24$s, SD $= 3.80$; $t(35) = -4.42$, $p < .001$)[4]. This suggests that regardless of training condition, infants noticed a change in object, looking more to the screen when this occurred.

As noted in our preregistration, we did not have *a priori* predictions that sex, age, or vocabulary size would explain variance in this study, but rather collected this information to better characterize the sample; see Supplementals for analyses confirming this prediction, and for results from the language background questionnaire. Participants were reported to understand 13.96 words on average (SD $= 13.76$), and produce 0.54 words (SD $= 1.06$).

---

[3] Within-talker variability: $M_{Same} = 4.89$, $M_{WordSwitch} = 6.13$. Between-talker variability: $M_{Same} = 6.81$, $M_{WordSwitch} = 6.24$.

[4] Within-talker variability: $M_{Same} = 4.89$, $M_{PictureSwitch} = 8.45$. Between-talker variability: $M_{Same} = 6.81$, $M_{PictureSwitch} = 10.04$.

Discussion

As predicted based on previous research, participants in the *No-Talker-Variability* condition dishabituated to all three types of switches: when the talker, word, or object changed. By contrast, participants in the two conditions featuring talker variability only increased their looking time to the *Picture-Switch* control, suggesting that while they accepted a previously learned word produced by a new talker, they also accepted a completely new word as a viable label for the trained object. We also found a difference in time to habituate across conditions, such that participants in the *Within-Talker-Variability* condition habituated faster than participants in the other two conditions. This result may suggest that within-talker variability could be easier to learn from, as it is most representative of infants' input (see Bulgarelli, Mielke, & Bergelson, in press). Before we move on to our next question of interest regarding how training with talker variability affects infants' sensitivity to mispronunciations in newly taught words, we first present a replication of the *No-Talker-Variability* condition which we conducted over Zoom (*Zoom Video Communications, Inc*, 2020) in response to the COVID-19 pandemic. Experiment 1b serves as a proof of concept that online data collection for a habituation study is comparable to data collection in the lab.

Experiment 1b

In Experiment 1b, we replicate the *No-Talker-Variability* condition in Experiment 1a with a new set of online data collection methods.

Methods

Participants. Our final sample was made up of 18 7- to 9-month-old infants (11 female, 7 male, M_{age} = 7.95 months). All participants were full term (40 +/- 3 weeks), monolingual (parents did not report >25% exposure to a language other than English), and had no history of hearing or vision problems. Participants were recruited from the broader area surrounding a university in the Southeastern United States and through childrenhelpingscience.com. Parents provided consent on behalf of themselves and their infants, and were compensated with a $5 Amazon gift card. 100% of the infants were reported by caretakers as White or Caucasian. Maternal education ranged from a high school degree to advanced degree (high school degree: n = 1; some trade school, professional training, or college: n = 1; associate or bachelor's degree: n = 9; advanced degree: n = 7). An additional 5 infants were excluded due to technical difficulties. Participants completed the experiment on a laptop or computer with a monitor size of 14" on average (ranging from 11 to 20"). The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board at Duke University.

Design. The design was the same as Experiment 1a, except that all participants were assigned to the *No-Talker-Variability* condition.

Stimuli. Stimuli were the same as those used in the Experiment 1a *No-Talker-Variability* condition.

Procedure. Instead of coming into the lab, participants joined a private Zoom room with the experimenter. After consent, infants sat in their caregiver's lap facing the computer or laptop in their own homes. The experimenter shared their screen such that all that was visible on the participants' screen was the experiment (e.g. participants could not see the video of themselves or of the experimenter, and the screen was in full screen mode). Parents were asked to not direct their infants' attention in any way and to keep the infant on their lap facing the computer if possible. In contrast to participation in the lab (Experiment 1a) parents were not asked to listen to cover music over headphones. As the sounds from the experiment were transmitted through the experimenter's computer speakers, the experimenter wore noise canceling headphones during the study to minimize access to the auditory stimuli (though it was impossible to be completely unaware of the auditory stimuli).

As we could not perfectly control the participants' distance to the monitor, prior to the warm up trials participants also saw a 9 point calibration video, which allowed the experimenter to gauge infants' looking pattern when looking at each edge of the screen. This made it easier to know when infants were looking off screen. Following the calibration video, the rest of the procedure was exactly as in Experiment 1a.

Results

Analysis Plan. For our main analysis, we conducted a mixed effects regression using lme4 to test whether the effects of test trial (Same vs. Talker Switch, Word Switch, and Picture Switch) differed by testing location: remote (over Zoom) or in the lab (using the data reported in Experiment 1a *No-Talker-Variability* condition). As above, we included subject random intercepts in the model[5]. Full model results, including Cohen's d can be found in Table 2, t and p values are also reported in text.

Reliability. Prior to reporting results, we wanted to make sure that reliability for live coding did not differ between Zoom studies and in-lab studies, especially since it was not possible for the experimenter to be completely unaware of the auditory stimuli presented through their computer for the Zoom participants. To evaluate this, an additional researcher, unaware of the experimental condition or trial order, was asked to code looking time offline for 5 Zoom participants and 5 in-lab participants. Offline coding was done in ELAN (Nijmegen: Max Planck Institute for Psycholinguistics, the Language Archive, n.d.), details can be found on OSF. In order to establish reliability, we computed correlations between looking times for trials (from habituation and test) coded live and offline. For in-lab studies, the correlation was $r = .94$, 95% CI $[.91, .96]$, $t(107) = 27.70$, $p < .001$, and for Zoom studies

[5] The model that also included the object-word pair and trial order random effect approached singularity, and thus these random effects were removed, as suggested by Barr et al., (2013).

Table 2

Fixed effects and Cohen's d for Experiment 1b model. '/' in predictor name indicates the specified #contrast (e.g. Same/TalkerSwitch compares looking time to Same vs TalkerSwitch trial); ':' indicates an interaction. SE is pooled for each predictor.

term	estimate	std.error	statistic	p.value	d
(Intercept)	7,951.31	402.93	19.73	<.001	NA
Same/TalkerSwitch	3,970.86	745.34	5.33	<.001	1.055
Same/WordSwitch	3,905.28	745.34	5.24	<.001	1.038
Same/PictureSwitch	3,822.56	745.34	5.13	<.001	1.016
Location	-498.62	402.93	-1.24	0.224	-0.424
Same/TalkerSwitch:Location	-657.14	745.34	-0.88	0.38	-0.175
Same/WordSwitch:Location	200.56	745.34	0.27	0.788	0.053
Same/PictureSwitch:Location	-1,197.11	745.34	-1.61	0.111	-0.318

it was $r = .95$, 95% CI [.92, .97], $t(78) = 26.44$, $p < .001$. These two high and similar correlations suggest that overall looking time across habituation and test trials was highly similar when coded online and when reliability-coded offline, both in the lab and over Zoom. As our analysis found that online coding for Zoom participants was highly accurate, we next report the looking-time results using the live-coded online data.

Habituation and Test Trial Results. Participants in the remote condition habituated after an average of 13.67 (SD = 4.79) trials. This did not differ significantly from the time to habituate in Experiment 1a's *No-Talker-Variability* condition (mean = 12.96, SD = 4.71), $t(33.88) = 0.61$, $p = .547$.

Results from the test trials are visualized in Figure 2 (1b panel), model output including estimates, standard errors and effect sizes are in Table 2, t and p values can be found in text. As above, we report main effects for each contrast first, followed by the interactions. Our model revealed an effect of trial, such that infants across both testing locations increased their looking time to the *Talker Switch* trial ($M_{TalkerSwitch}$ = 9.00s, SD = 3.46), the *Word Switch* trial ($M_{WordSwitch}$ = 8.93s, SD = 4.26) and the *Picture Switch* trial ($M_{PictureSwitch}$ = 8.85s, SD = 4.28) relative to the *Same* trial (M_{Same} = 5.03s, SD = 2.38; Same vs. Talker Switch: t = 5.33, p < .001; Same vs. Word Switch: t = 5.24, p < .001; Same vs. Picture Switch: t = 5.13, p < .001).

The effect of location (online vs. Zoom) was not significant, t = -1.24, p = .224, and neither were any of the interactions (Same vs. Talker Switch by location: t = -0.88, p = .380; Same vs. Word Switch by location: t = 0.27, p = .788; Same vs. Picture Switch by location: t = -1.61, p = .111). These results suggest that the pattern of looking time and the length of looking did not differ across participants in the lab vs. over Zoom.

Discussion

The results of the participants collected over Zoom fully replicate the pattern of results seen from participants in the *No-Talker-Variability* condition from Experiment 1a that was collected in the lab. This is itself an important contribution, as several parameters varied across these testing locations. The most notable differences between lab and Zoom to us were that the experimenter could not be completely unaware of the stimuli presented to participants, that caregivers were not asked to listen to masking music, and that the size of the monitor or distance to the monitor were not controlled in participants' homes as they were in the lab. Nevertheless, Exps. 1a and 1b rendered identical patterns of results. This lets us more confidently move on to our originally designed Experiment 2, conducted mostly online.

Experiment 2

In Experiment 2, we ask whether talker variability during habituation can help infants reject mispronunciations of a newly-trained word. On the one hand, as infants in the talker variability conditions in Experiment 1a did not dishabituate when the word changed entirely (heard ‘neem’ after being trained with ‘lof’), it would be surprising if infants rejected a more subtle change in vowel (‘noom’ instead of ‘neem’) when paired with the trained object. Nonetheless, it is possible that at this early stage of word recognition, talker variability is particularly relevant for distinguishing between minimal pairs (Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009). Thus, instead of the *Talker Switch* trial used in Experiment 1, Experiment 2 uses a *Mispronunciation (MP) Switch* where the vowel of the trained word changes, but the object and talker remain the same. We originally preregistered this Experiment along with Experiment 1a (https://osf.io/acrsp), and then amended our preregistration to reflect that in Experiment 1a, participants did not dishabituate to the Word Switch, and thus may also fail to dishabituate to a Mispronunciation Switch (https://osf.io/73wbq). A power analysis revealed that a sample size of 18 participants per condition would be appropriate, as detailed in Experiment 1a.

Methods

Participants. Our final sample was made up of 54 7- to 9-month-old infants (28 female, 26 male, M_{age} = 7.70 months). All participants were full term (40 +/- 3 weeks), monolingual (parents did not report >25% exposure to a language other than English), and had no history of hearing or vision problems. Eight participants were tested in the lab prior to the COVID-19 pandemic; the remainder were tested online. Participants were recruited from the broader area surrounding a university in the Southeastern United States and through childrenhelpingscience.com. Parents provided consent on behalf of themselves and their infants, and were compensated with mileage reimbursement and a child-focused thank you gift (in the lab) or a $5 Amazon gift card (online).

We report race breakdown based on testing location. For participants tested in the lab, 88% were White or Caucasian, and 12% were Asian. For participants tested online, 91% of the infants were White or Caucasian, 2% were Asian, and 7.00% identified as other or multiracial. Maternal education ranged from a high school degree to advanced degree (high school degree: n = 1; some trade school, professional training, or college: n = 1; associate or bachelor's degree: n = 14; advanced degree: n = 35). An additional 12 infants were excluded; 5 due to technical difficulties, 4 due to not meeting looking time criteria (during habituation or at test), 1 due to not making it through the entire experiment, 1 due to parental interference, and 1 due to experimenter error. Participants over Zoom completed the experiment on a laptop or computer with a monitor size of 15" on average (ranging from 12 to 34"). The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board at Duke University.

Design. As in Experiment 1a, participants were habituated to a single word-object pair in one of three talker variability conditions. The warm up trials and habituation phase, as well as the *Same* trial, the *Word Switch* and *Picture Switch* were identical to Experiment 1a. The only change was that the critical *Talker Switch* test trial was replaced with a *Mispronunciation (MP) Switch* test trial.

Stimuli. Stimuli were the same as those used in the Experiment 1a, with the addition of a two feature mispronunciation to the vowel of the non-word (changing the frontness and roundness of the vowel). The mispronunciation of 'neem' was 'noom' and the mispronunciation of 'lof' was 'lef.' These were recorded in the same way as the rest of the stimuli in Experiment 1a, described above.

Caregiver questionnaires. As in Experiment 1a, caregivers filled out a CDI, a language exposure survey, and a demographics questionnaire. Results from the CDI and language exposure survey can be found in Supplementals.

Procedure. Eight participants were tested in the lab prior to the COVID-19 pandemic, and thus the procedure for them was identical to that described in Experiment 1a. The remainder of participants were tested over Zoom, and thus the procedure for them was identical to that described in Experiment 1b. For all participants, the warm-up trials and habituation phase were identical to Experiment 1a.

Test trials. After the habituation phase, infants were advanced to the test phase, which consisted of four test trials: a *Same* trial and three *Switch trials*, each lasting up to 14s. The *Same* test trial repeated a token used during habituation. The critical *Mispronunciation (MP) Switch* test trial repeated a single token of a mispronounced version of the habituated word where the vowel changed, spoken by a talker heard during habituation (e.g. “lef” for “lof”); this tested infants’ ability to reject an incorrect pronunciation of the learned word. As in Experiment 1a, the *Word Switch* and *Picture Switch* control trials queried whether infants detected when each component of the word-object link was broken. All test trials in Experiment 2 featured a talker from the habituation phase.

Results

Analysis Plan. The analysis plan for Experiment 2 was identical to that used in Experiment 1a, except that the contrast comparing the Talker Switch to the Same trial was replaced with one comparing the Mispronunciation Switch to the Same trial. To account for possible stimuli idiosyncrasies, we included random intercepts for word-object pair (which by design also includes talker), as well as by-Subject random intercepts in the model[6]. As above,

[6] The model including the test order random effect approached singularity, and thus the random effect of order was removed as suggested by Barr et al., (2013).

we do not report omnibus effects for each variable, and instead report results for our specific contrasts of interest.

Habituation and Test Trial Results. Across all three habituation conditions, infants habituated after an average of 12.57 (SD = 5.89) trials. This did not differ by habituation condition, $F(2, 51) = 0.49$, $MSE = 35.35$, $p = .617$.

Results for the test trials are visualized in Figure 3; full model results, including Cohen's d can be found in Table 3, t and p values are also reported in text. As for Exp. 1a, we report main effects for each contrast first, followed by the interactions.

There were no main effects of habituation condition: looking time did not differ *overall* between the *No-Talker-Variability* condition and the two talker variability conditions (t = 1.22, p = .229), nor between the *Within-Talker-Variability* and the *Between-Talker-Variability* conditions (t = -0.66, p =.514).

We did find a significant effect of trial, such that infants across all conditions increased their looking time to the control trials, i.e. the *Word Switch* trial ($M_{\text{WordSwitch}}$ = 6.97s, SD = 4.47) and the *Picture Switch* trial ($M_{\text{PictureSwitch}}$ = 9.21s, SD = 4.00) relative to the *Same* trial (M_{Same} = 5.34s, SD = 3.42); Same vs. Word Switch: $t = 2.58$, $p = .011$, Same vs. Picture Switch: $t = 6.15$, $p < .001$). However, looking time to the critical *Mispronunciation Switch* did not differ from looking time to the *Same* trial (M_{MPSwitch} = 6.16s, SD = 3.58), $t = 1.30$, $p = .196$.

We next turn to the interactions between the trial type contrasts and the habituation condition contrasts. The interaction between the trial type contrasts comparing the *WordSwitch* trial vs. the *Same* trial and the contrast comparing the within- and between-talker variability was not significant, $t = 1.77$, $p = .079$. Given this, we do not interpret this result any further[7].

[7] For full transparency for interested readers given the marginal (but not significant) p-value, we provide the relevant t-test and condition means. Namely, while looking time to the *Word Switch* trial was significantly

Table 3

Fixed effects for Experiment 1b model, as well as Cohen's d. '/' in predictor name indicates the specified contrast (e.g. Same/MPSwitch compares looking time to Same vs MPSwitch trial); ':' indicates an interaction. SE is pooled for each predictor.

term	estimate	std.error	statistic	p.value	d
(Intercept)	6,920.72	590.49	11.72	0.054	NA
Same/MPSwitch	817.96	629.37	1.30	0.196	0.210
Same/WordSwitch	1,621.54	629.37	2.58	0.011	0.417
Same/PictureSwitch	3,869.37	629.37	6.15	<.001	0.994
NoVariability/TalkerVariability	918.84	755.14	1.22	0.229	0.344
WithinTalker/BetweenTalker	-286.58	435.98	-0.66	0.514	-0.186
Same/MPSwitch:NoVariability/TalkerVariability	934.14	1,335.09	0.70	0.485	0.113
Same/WordSwitch:NoVariability/TalkerVariability	1,055.44	1,335.09	0.79	0.43	0.128
Same/PictureSwitch:NoVariability/TalkerVariability	-1,123.89	1,335.09	-0.84	0.401	-0.136
Same/MPSwitch:WithinTalker/BetweenTalker	257.08	770.82	0.33	0.739	0.054
Same/WordSwitch:WithinTalker/BetweenTalker	1,363.11	770.82	1.77	0.079	0.286
Same/PictureSwitch:WithinTalker/BetweenTalker	640.72	770.82	0.83	0.407	0.134

Similarly, the interactions comparing looking time to the critical *Mispronunciation Switch* and the control *Picture Switch* test trials to the *Same* test trial across the two conditions featuring talker variability were not significant (all ps > 0.41). This suggests that looking times for these comparisons did not vary as a function of between- vs. within-talker variability during habituation.

higher than to the *Same* trial in the *Between-Talker-Variability* condition ($M_{WordSwitch} = 6.89$s (SD = 4.28); $M_{Same} = 4.26$s (SD = 2.83)); $t(17) = -2.68$, $p = .016$, this was not the case in the *Within-Talker-Variability* condition, ($M_{WordSwitch} = 5.87$s (SD = 4.34); $M_{Same} = 5.96$s (SD = 4.04); $t(17) = 0.08$, $p = .937$).

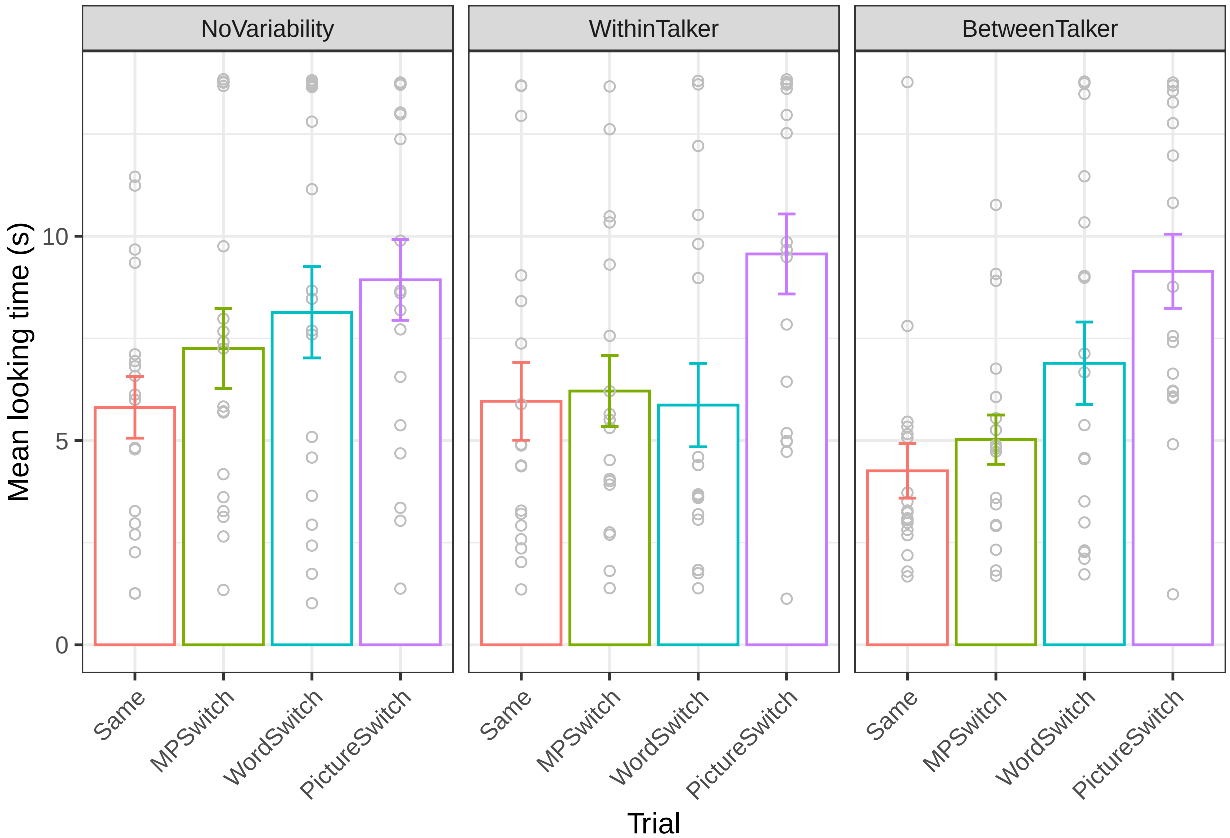

Figure 3. Results for Experiment 2. Bars reflect mean looking time (y-axis) across test trials for participants in all three conditions (x-axis). Circles indicate individual data points, error bars reflect standard error.

Unlike in Experiment 1, the interactions between the trial type contrasts and the contrast comparing the *No-Talker-Variability* condition to the two conditions featuring talker variability were not significant: *Mispronunciation Switch* trial vs. *Same* trial, t = 0.70, p = .485; *WordSwitch* vs. *Same* trial, t = 0.79, p = .430; *PictureSwitch* vs. *Same* trial, t = -0.84, p = .401. This suggests that looking time patterns across test trials did not differ *overall* across habituation conditions, as a function of talker variability.

As noted in our preregistration, we did not have *a priori* predictions that sex, age, or vocabulary size would explain variance in this study, but rather collected this information to better characterize the sample; see Supplementals for analyses confirming this prediction. Participants were reported to understand 10.42 words on average (SD = 10.85), and produce

0.35 words (SD = 0.96).

Patterns across experiments. In an exploratory analysis, we pooled the data from Experiments 1 and 2 to further consider two results that varied across experiments. Namely, we explored whether the number of trials to habituate differed by condition, and whether looking time to the *Word Switch* control differed from the *Same* test trials across these same training conditions. In brief, we find that across experiments, (1) participants habituated faster in the *Within-Talker-Variability* condition, relative to the other two habituation conditions; and (2) that talker variability during training led infants to incorrectly accept a completely novel word as the label for the trained object (e.g. "lof" as a label for what they were trained was a 'neem'). We underscore the exploratory nature of these analyses, and suggest they should be replicated in future research to ascertain their reliability. Details of these analyses are available in the Supplementals.

Discussion

In Experiment 2, we tested whether infants trained on a word-object pairing with or without talker variability would dishabituate if they heard the vowel in the trained word mispronounced (e.g. lef for lof). We found that regardless of habituation condition, participants did not dishabituate to the Mispronunciation Switch trial, suggesting that 8-month-old infants do not notice when newly-learned words are mispronounced in this context. Furthermore, we found that as in Experiment 1a, participants in all three training conditions dishabituated to the Picture Switch control, noticing when a completely novel object was paired with the habituated word. The results for the Word Switch control trial fell between these two patterns. That is, how infants treated a completely phonetically different word paired with a trained object varied in a complicated way as a function of the talker variability they were trained with. We return to this in the general discussion.

General Discussion

Across two experiments, we asked whether manipulating talker variability while teaching 8-month-olds a single new word-object pairing would lead them to an adult-like conclusion: that a change in talker does not change a word's identity, but that a change to a single phoneme does. To ask this, we first habituated infants to a new word-object link with or without talker variability. We then presented them with an instance of the word and object they were trained with on one trial, and altered how the word sounded relative to their training on another, either with a new talker (Exp. 1a; *Talker Switch*) or a mispronunciation to the central vowel (Exp. 2; *Mispronunciation Switch*). We also included control trials checking that infants had made the word-object link in the first place by changing the word or depicted object altogether (*Word Switch* and *Picture Switch*, both Experiments). The premise of this manipulation is that infants' looking time serves as a proxy for whether they find the changes that we made to be critical for the word-object link (Stager & Werker, 1997). That is, by hypothesis, infants look longer to changes that break this link vs. those that do not.

Consistent with our predictions, infants in the *No-Talker-Variability* condition dishabituated to the novel talker (on the *Talker Switch*, Experiment 1a and 1b), but did not dishabituate to the mispronunciation (on the *MP Switch*, Experiment 2). They also exhibited the early hallmarks of word learning, dishabituating to both the *Word Switch* and *Picture Switch* control trials in Experiment 1a, 1b, and 2, replicating Werker, Cohen, Lloyd, and Stager (1998). In turn, these results suggest that while this very acoustically-narrow training experience (i.e. a single word token) led infants to correctly reject some cases in which the word-object link was broken, it also led them to incorrectly *reject* new talkers, and incorrectly *accept* mispronunciations. These results are consistent with prior work (Houston & Jusczyk, 2000; Swingley, 2005), and show that initial word-object links after training with no talker variability in 8-month-olds are not yet adult-like.

Might talker variability in training help infants form more appropriate word-object links? While we correctly predicted the pattern of results in the *No-Talker-Variability* condition, our predictions for participants in the *Within-* and *Between-Talker-Variability* conditions were only partially borne out. We found that consistent with appropriate bounds on word-object links, infants in these talker variability conditions did not dishabituate when they heard the newly trained word produced by a new talker in Experiment 1a (*Talker Switch*), but did dishabituate when the object changed (*Picture Switch* control trial) in Exp. 1a and 2. On the other hand, they also failed to dishabituate to the mispronunciation in Experiment 2 (*MP Switch*), and even the fully new word (*Word Switch* control trial) in both experiments. The divergence in the patterns between the no talker variability condition and the two conditions featuring talker variability suggests that talker variability during training altered how infants treated sound-based changes to the word-object link. However, this training seems to have led infants too far in this direction: the results suggest infants accepted changes that should have indicated a break in the trained word-object link (i.e. mispronunciation of the key vowel and a fully different word). This behavior too is not yet adult-like.

Word-object link formation

A fair concern raised by our results is whether infants' lack of dishabituation to the *Word Switch* control trial in the *Within-* and *Between-Talker-Variability* conditions indicates that they failed to learn the word-object link at all. To address this possibility, it's helpful to first consider how infants would have behaved if they attended to only one modality of the input, not attempting to link the word and object together. Focusing on the auditory modality first, recent results find that in the absence of a visual referent, 7.5-month-olds trained on /bIm/ with 1 or 4 talkers dishabituate upon hearing /pIm/ (Quam, Clough, Knight, & Gerken, 2020). Similarly, Von Holzen and Nazzi (2020) find that 8 month old infants notice vowel mispronunciations of their own names, suggesting that even the type of mispronunciation used here is salient at this age. This suggests infants at the age tested here can discriminate minimal pairs of sounds (albeit different phonemic changes), even when presented with multiple talkers during habituation.

Indeed, auditory discrimination was suggested as the reason 8-month-olds succeeded on a single-object switch using minimal pairs in the original Stager and Werker (1997) study. That is, Stager and Werker (1997) argued that while 14-month-olds failed to detect the difference in a single-object switch using minimal pairs because they were engaged in word-object mapping, 8-month-olds detected this same difference because they were treating it as a sound discrimination task. By this logic, our 8-month-olds across all conditions are behaving like Stager and Werker (1997)'s 14-month-olds, i.e. treating the *Mispronunciation Switch* like the *Same* trial. If infants were simply engaging in a sound discrimination task, we'd expect them to dishabituate to all types of auditory changes that they can detect. Instead, infants in all conditions here did not dishabituate to the *Mispronunciation Switch* and infants in the conditions featuring talker variability also did not dishabituate to the *Word Switch* control trial. This pattern of results suggests that the current task went above and beyond a simple sound discrimination task, possibly due to the presence of warm-up trials which established the referential nature of the task, which has been shown to help 14-month-olds (Fennell & Waxman, 2010). Here too, these warm-up trials may have edged the 8-month-olds towards a word-object mapping task as well.

Another way in which infants in the talker variability conditions could have failed to form the word-object link would be if they focused solely on the visual object and ignored the auditory input altogether. While in principle possible, we find this unlikely, based on our habituation analysis across experiments and on infants' experiences in everyday life. If infants attended only to the visual information, we would have expected to see no differences in time to habituate across conditions that varied only in the auditory input. Instead, our exploratory habituation analysis found that across experiments, infants habituated faster to the *Within-Talker Variability* condition than the other two conditions, which seems difficult to explain if infants are ignoring the auditory input (see Supplementals).

Relatedly, talker variability is rampant in infants' daily lives. While infants from a similar background to the current sample generally hear most of their noun input from one talker, they also generally hear many talkers a day (Bergelson & Aslin, 2017; Bulgarelli, Mielke, & Bergelson, in press). In fact, toys and media are likely the only sources that provide highly consistently instances of words. While the prevalence of such electronic and consistent tokens varies across households, it likely makes up a very small proportion of the input on average, e.g. only 5% of nouns were produced by electronic sources in a corpus of daylong recordings from 44 infants from a similar background to those tested here (see Bulgarelli & Bergelson, 2019). Indeed, the variability infants were exposed to here deliberately mimicked real-world variability[8]: rather than exposing infants to stimuli parametrically varying one acoustic property at a time, we deliberately varied many properties simultaneously (duration, prosody, contour, etc), using natural speech tokens more akin to infants' daily experiences. Thus, given infants' consistent experience learning from variable tokens of words, it would be surprising if they chose to not attend to the auditory input in our experiments altogether.

Instead, we conclude that infants in all conditions likely attended to both the auditory and visual information presented during training. Our interpretation of the results is that rather than only engaging in a word-object learning task in the No-Talker Variability condition, infants did so across conditions. However, the addition of talker variability during habituation led to differences in what was learned. That is, training with talker variability may alter what infants attend to, allowing them to learn how the surface features of the word can vary (and thus accept the trained word produced by a new talker), but making it more difficult to learn which sound-based changes break the word-object link (failing to reject the mispronunciation and the word changes).

[8] This holds within the current sample's cultural context; whether talker variability manifests differently cross-culturally is an open question.

Why might this be? One possibility is that the increased complexity of the learning task as a result of talker variability can be beneficial for some aspects of processing (i.e. generalizing to new talkers) but challenging for others (i.e. not generalizing to new words). This is consistent with previous research which has found that acoustic variability can benefit generalization (e.g. Singh, 2008) and invariance detection on one hand (e.g. Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009), but can also potentially overwhelm learners (e.g. Quam, Knight, & Gerken, 2017) or slow down the process of learning on the other (e.g. Van Heugten & Johnson, 2017; see Quam & Creel, 2021 for a review of variability on aspects of language development; and see Bulgarelli & Weiss, 2021 for relevant work with adults). Thus, the current results may reveal evidence for both facilitation and inhibition of processing and/or learning within a single task.

Another possibility is that training with variability broadened learners' expectations about how future input could sound, consistent with the *general expansion mechanism* proposed by Schmale and colleagues for accent accommodation (Schmale, Cristia, & Seidl, 2012; Schmale, Seidl, & Cristia, 2015). In Schmale et al's studies, toddlers exposed to either multiple talkers producing speech or silent videos of multiple individuals prior to a word learning task went on to accommodate accent variability for newly trained words, while toddlers who were not exposed to variability prior to learning did not (Schmale, Cristia, & Seidl, 2012; Schmale, Seidl, & Cristia, 2015). Here too, training with talker variability led infants to accept any auditory change to the word-object link: a change in talker and a change in word. Thus, even before age one, infants can employ this *general expansion mechanism* to accommodate talker information, allowing them to learn how the surface features of the word can vary. However, employment of this mechanism might be initially immature, reflected by infants' incorrect extensions to large sound-based changes that break the word-object link (i.e. failing to reject the mispronunciation and the word changes). This fits nicely with the proposal set forth by Schmale, Cristia, and Seidl (2012); suggesting that while this general expansion mechanism can be useful, it could also lead to accepting inappropriate changes (in their proposal, in accented speech) that are not supported by evidence in the input (i.e. that neem is a viable token of lof).

Of course, participants in the *No-Talker-Variability* condition also struggled with appropriately noticing what kinds of more subtle changes break the word-object link, since they accepted the mispronunciation but rejected a new talker. The adult-like pattern is to (1) consider talker changes irrelevant to word identity, i.e. treat the trained word said by the new talker just like the same word said by the familiar talker; and (2) to consider a mispronunciation and a new word a poor fit for the trained object. Infants in the *No-Variability* training condition failed to do (1), dishabituating to the *Talker Switch* trial, and failed to do part of (2), not noticing when the word was mispronounced. Infants in the two talker variability conditions succeeded at (1), but failed to do (2): they failed to dishabituate when the word that went with the trained object changed a little, and a lot. Clearly, 8-month-olds are not yet adult-like in their early word learning, though intriguingly, the variability in their training leads to different patterns of behavior.

Collectively, our results show that by 8 months of age, infants' process of forming a new word-object link (in the lab) is shaped by brief exposure to acoustically variable stimuli. Hearing words more variably changes what infants attend to and the concomitant word representations they form (see also Singh (2008); Van Heugten and Johnson (2017)). As mentioned above, the kinds of within- and between-talker variability tested here are akin to what infants from a similar background are exposed to in their everyday lives, i.e. many tokens of words from the same talker, and ~6 distinct talkers a day (Bergelson & Aslin, 2017; Bulgarelli, Mielke, & Bergelson, in press); this contrasts with approaches that expose infants to unfamiliar inputs, such as novel accents (Potter & Saffran, 2017). While the precise circumstances in which variability may facilitate or inhibit learning remain unsettled, infants' own experience with variability may provide some insight. That is, the timeline of learning to appropriately interpret variability in talker, affect and accent may itself be influenced by early and extensive exposure to such variability. Further, the effect of variability on learning may also depend on whether the variability is non-contrastive and signals invariance, and learners should therefore generalize across it, or whether it signals contrastive dimensions of the input and should be attended to (see Apfelbaum & McMurray, 2011; Gogate & Hollich, 2010), such as in the case of multi-dialectal or multi-lingual environments. Thus, the age at which infants can harness variability to appropriately expand their expectations regarding future input is an exciting and open question.

Within- and between- talker variability

Even though the two types of talker-variability used here provided different acoustic information (see Galle, Apfelbaum, & McMurray, 2015), the pattern of looking times on the test trials did not differ depending on whether infants heard *Within* or *Between* talker variability. This is consistent with previous research using the minimal-pair switch task in which both within- and between-talker variability affected learning equivalently (Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009; for effect sizes see Tsui, Byers-Heinlein, & Fennell, 2019). This suggests that 8-month-old infants (tested here) and 14-month-old infants (in Galle, Apfelbaum, & McMurray, 2015; Rost & McMurray, 2009) treat talker variability stemming from a single talker or from multiple talkers similarly, at least for the purposes of initial word-object links and word recognition.

Task considerations

While the two-word switch task has been widely used to test word learning in one-year-olds (see Tsui, Byers-Heinlein, & Fennell, 2019 for a meta-analysis); the one-word switch task is less common. Here, we demonstrated that the single word switch task can be used to probe early aspects of word learning in 8-month-olds. However, our results highlighted the intrinsic limitations of the single-word switch task. Namely, by dint of only teaching one word-object link, there is a limited set of parameters that can be varied to query exactly what infants learned. That is, we could not test whether infants had learned the word-object link without introducing untrained novel objects or words, in contrast to the traditional two-word switch task. Another option for future work might be to incorporate familiar words or objects (e.g. the label 'dog' paired with a newly trained object or a picture of dog paired with the trained word), though this has its own interpretive challenges. Given that infants at this age have already begun understanding common nouns (see e.g. Bergelson & Aslin, 2017; Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012), understanding how we can teach new words and query learning in the lab at young ages is important for uncovering how this process unfolds in everyday life.

Our results also show that the one-word switch task can readily be adapted for online data collection. However, it's worth noting that our online samples were less racially and ethnically diverse than our lab-based sample. This could be due to our recruitment protocol during COVID-19, which only allowed us to contact families that had signed themselves up to participate (as opposed to our typically broader community-based recruitment approach), or to the need to have a computer or laptop as well as an internet connection to participate remotely. While online data collection has the potential to reach a broader audience relative to the participant pool that is willing and able to come to campus to participate, it still presents some inherent recruitment challenges.

Conclusion and Future directions

Taken together, our results suggest that talker variability influences newly forged word-object links in eight-month-olds. We find that in a controlled lab study, both within- and between-talker variability change how word learning unfolds relative to exposure to a new word without talker variability. This provides first steps in understanding how our youngest word-learners leverage 'relevant' and 'irrelevant' acoustic variability to eventually build properly-constrained representations of words within their nascent lexicons. Nonetheless, just how variability between and within talkers gets consolidated and codified into appropriately specific representations of common words—both in the lab and in daily life—remains an open question for future research. We invite and look forward to further work establishing the conditions under which infants learn to treat talker- and phoneme-based differences in adult-like ways during word learning.

References

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*(6), 1105–1138. https://doi.org/10.1111/j.1551-6709.2011.01181.x

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Bates, D., & Maechler, M. (2019). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, 201712966. https://doi.org/10.1073/pnas.1712966114

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, *18*(4), 587–598. https://doi.org/10.1111/desc.12242

Bulgarelli, F., & Bergelson, E. (2019). Look who's talking: A comparison of automated and human-generated speaker tags in naturalistic day-long recordings. *Behavioral Research Methods*, *1-13*. https://doi.org/10.3758/s13428-019-01265-7

Bulgarelli, F., Mielke, J., & Bergelson, E. (in press). Quantifying talker variability in North-American infants' daily input. https://doi.org/10.1111/cogs.13075

Bulgarelli, F., & Weiss, D. J. (2021). Desirable Difficulties in Language Learning? How Talker Variability Impacts Artificial Grammar Learning. *Language Learning*, 1–37. https://doi.org/10.1111/lang.12464

Casasola, M., & Cohen, L. B. (2000). Infants' association of linguistic labels with causal actions. *Developmental Psychology*, *36*(2), 155–168. https://doi.org/10.1037/0012-1649.36.2.155

Fennell, C. T. (2012). Object Familiarity Enhances Infants' Use of Phonetic Detail in Novel Words. *Infancy*, *17*(3), 339–353. https://doi.org/10.1111/j.1532-7078.2011.00080.x

Fennell, C. T., & Waxman, S. R. (2010). What Paradox ? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning. *Child Development*, *81*(5), 1376–1383. https://doi.org/10.1111/j.1467-8624.2010.01479.x

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). *Variability in Early Communicative Development* (No. 5; Vol. 59, pp. 1–185). Retrieved from https://www.jstor.org/stable/pdf/1166093.pdf?refreqid=excelsior%7B/%%7D3A28b49b2d69ee2a5cd880edbb428aad5b

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (Second). Thousand Oaks CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Fox, J., Weisberg, S., & Price, B. (2018). *carData: Companion to applied regression data sets*. Retrieved from https://CRAN.R-project.org/package=carData

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2015). *The Role of Single Talker Acoustic Variation in Early Word Learning The Role of Single Talker Acoustic*

Variation in Early Word Learning. (December). https://doi.org/10.1080/15475441.2014.895249

Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: a perceptual gateway to language development. *Psychological Review*, *117*(2), 496–516. https://doi.org/10.1037/a0019049

Henry, L., & Wickham, H. (2019). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Hlavac, M. (2018). *Stargazer: Well-formatted regression and summary statistics tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). Retrieved from https://CRAN.R-project.org/package=stargazer

Hohle, B., Fritzsche, T., Meb, K., Philipp, M., & Gafos, A. (2020). Only the right noise? Effects of phonetic and visual input variability on 14-month-olds' minimal pair word learning. *Developmental Science*, 0–2. https://doi.org/10.1111/desc.12950

Houston, D. M. (1999). *The role of talker variability in infant word representations (Unpublished doctoral dissertation)*. (PhD thesis). Johns Hopkins University, Baltimore, MD.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. https://doi.org/10.1037/0096-1523.26.5.1570

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. W. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, *43*, 253–291. https://doi.org/10.1016/0010-0277(92)90014-9

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Liberman, A. M., Coopers, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6). https://doi.org/10.1037/h0046234

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

Müller, K., & Wickham, H. (2019). *Tibble: Simple data frames.* Retrieved from https://CRAN.R-project.org/package=tibble

Nijmegen: Max Planck Institute for Psycholinguistics, the Language Archive. (n.d.). ELAN (Version 5.9). Retrieved from https://archive.mpi.nl/tla/elan

Oakes, L. M., Sperka, D., DeBolt, M. C., & Cantrell, L. M. (2019). Habit2: A stand-alone software solution for presenting stimuli and recording infant looking times in order to study infant development. *Behavior Research Methods*, *51*(5), 1943–1952. https://doi.org/10.3758/s13428-019-01244-y

Polka, L., & Werker, J. F. (1994). Developmental Changes in Perception of Nonnative Vowel Contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Potter, C. E., & Saffran, J. R. (2017). Exposure to multiple accents supports infants ' understanding of novel accents. *Cognition*, *166*, 67–72. https://doi.org/10.1016/j.cognition.2017.05.031

Quam, C., Clough, L., Knight, S., & Gerken, L. A. (2020). Infants' discrimination of consonant contrasts in the presence and absence of talker variability. *Infancy*, (December 2019), 1–20. https://doi.org/10.1111/infa.12371

Quam, C., & Creel, S. C. (2021). Impacts of acoustic-phonetic variability on perceptual development for spoken language : A review. *Wiley Interdisciplinary Reviews: Cognitive Science*, (September 2020), 1–21. https://doi.org/10.1002/wcs.1558

Quam, C., Knight, S., & Gerken, L. (2017). *The Distribution of Talker Variability Impacts Infants ' Word Learning*. https://doi.org/10.5334/labphon.25

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

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

Robinson, D., & Hayes, A. (2019). *Broom: Convert statistical analysis objects into tidy tibbles*. Retrieved from https://CRAN.R-project.org/package=broom

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

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

RStudio Team. (2019). *RStudio: Integrated development environment for r*. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Ryalls, B. O., & Pisoni, D. B. (1997). The Effect of Talker Variability on Word Recognition in Preschool Children. *Dev Psychol.*, *33*(3), 441–452.

https://doi.org/10.1037/0012-1649.33.3.441

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

Schmale, R., & Seidl, A. (2009). The role of variability in voice and foreign accent in the development of early word representations. *Developmental Science*, *70*(1), 0718. https://doi.org/10.1111/j.1467-7687.2009.00809.x

Schmale, R., Seidl, A., & Cristia, A. (2015). Mechanisms underlying accent accommodation in early word learning: evidence for general expansion. *Developmental Science*, *18*(4), 664–670. https://doi.org/10.1111/desc.12244

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*(2), 833–870. https://doi.org/10.1016/j.cognition.2007.05.002

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*(2), 173–189. https://doi.org/10.1016/j.jml.2004.04.004

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Letters to Nature*, 381–383. https://doi.org/10.1038/41102

Swingley, D. (2005). 11-Month-Olds' Knowledge of How Familiar Words Sound. *Developmental Science*, *8*(5), 432–443. https://doi.org/10.1111/j.1467-7687.2005.00432.x

Tincoff, R., & Jusczyk, P. W. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological Science*, *10*(2), 172–175. https://doi.org/10.1111/1467-9280.00127

Tincoff, R., & Jusczyk, P. W. (2012). Six-Month-Olds Comprehend Words That Refer to Parts of the Body. *Infancy*, *17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Tsui, A. S. M., Byers-Heinlein, K., & Fennell, C. T. (2019). Associative word learning in infancy: A meta-analysis of the Switch task. *Developmental Psychology*, *55*(5), 934–950. https://doi.org/10.1037/dev0000699

Van Heugten, M., & Johnson, E. K. (2017). Input matters: Multi-accent language exposure affects word form recognition in infancy. *The Journal of the Acoustical Society of America*, *142*(2), EL196–EL200. https://doi.org/10.1121/1.4997604

Von Holzen, K., & Nazzi, T. (2020). Emergence of a consonant bias during the first year of life: New evidence from own-name recognition. *Infancy*, *25*(3), 319–346. https://doi.org/10.1111/infa.12331

Werker, J. F., Cohen, L. B., Lloyd, V. L., & Stager, C. L. (1998). *Acquisition of Word-Object Associations by 14-Month-Old Infants*. *34*(6), 1289–1309.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H. (2019). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2019). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Francois, R. (2018). *Readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.name/knitr/

Zhu, H. (2019). *kableExtra: Construct complex table with 'kable' and pipe syntax*. Retrieved from https://CRAN.R-project.org/package=kableExtra

Zoom video communications, inc (Version 5.7.6). (2020). Retrieved from https://zoom.us/