Talker variability is not always the right noise: 14 month olds struggle to learn dissimilar word-object pairs under talker variability conditions

Federica Bulgarelli[1,2], **Elika Bergelson**[2]

[1]University at Buffalo

[2]Duke University

Abstract

Seminal work by Stager & Werker (1997) finds that 14-month-olds can rapidly learn two word-object pairings if the words are distinct (e.g. "neem" and "lif") but not similar (e.g. the minimal pair "bih" and "dih"). More recently, studies have found that adding talker variability during exposure to new word-object pairs lets 14-month-olds succeed on the more challenging minimal pair task, presumably due to talker variability highlighting the "relevant" consistencies between the similar words (Rost & McMurray, 2009; Galle et al., 2015; Hohle et al., 2020). It remains an open question, however, whether talker variability would be similarly useful for learning new word-object pairings when the words themselves are already distinct, or whether instead this extra variability may extinguish learning due to increased task demands. We find evidence for the latter. Namely, in our sample of 54 English-learning 14-month-olds, training infants on two word-object pairings (e.g. "neem" with a dog toy and "lof" with a kitchen tool) only led them to notice when the words and objects were switched if they were trained with single-speaker identical word tokens. When the training featured talker variability (from one or multiple talkers) infants failed to learn the pairings. We suggest that when talker variability is not necessary to highlight the invariant differences between similar words, it may actually increase task difficulty, making it harder for infants to determine what to attend to in the earliest phases of word learning.

Keywords

talker variability; acoustic variability; switch task; word-object mapping; infant word learning; phonological development

Introduction

In order to build their vocabulary, infants must integrate sequences of speech sounds and referents in the world, building what ultimately become stable word-meaning pairs. Even by

6 months of age, infants know the meanings of some common words, such as body parts and foods (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012); more robust knowledge (dubbed the "comprehension boost") emerges just around the first birthday (Bergelson, 2020). But how does this learning come about and what influences whether words are easier or harder to learn? One common method for studying the earliest phases of this process is through associative learning tasks, such as the Switch task.

In a seminal demonstration of infants' associative learning skills, 14-month-old infants were habituated to two novel word-object pairs (e.g. novel-object-1 with the word "neem" and novel-object-2 with the word "lif"; Stager and Werker (1997)). Learning of the word-object pairs was tested by comparing looking time to a *Same* trial (featuring the same pairings infants were trained on) versus looking time on a *Switch* trial (swapping the pairing infants were trained on, e.g. "lif" with novel-object-1). When these words were distinct sounding (e.g "neem" and "lif"), infants increased their looking time to the *Switch* trial, which is taken as evidence that infants had formed the word-object link, and were surprised when it was broken. When these words were similar sounding minimal pairs (e.g. "bih" and "dih"), infants did not increase their looking time to the *Switch* trial (see also Hoehle, Fritzsche, Mess, Philipp, & Gafos, 2020; Pater, Stager, & Werker, 2004; Rost & McMurray, 2009). That is, even though infants can distinguish the two similar sounding words in the absence of a visual referent (Stager & Werker, 1997), they were not able to map these similar sounding words to distinct novel objects.

Subsequent research has investigated the conditions under which 14-month-olds can learn novel minimal pairs. One line of research has shown that infants can succeed when they are provided with additional scaffolding highlighting the referential nature of the task (Fennell, 2012; Fennell & Waxman, 2010). For example, Fennell and Waxman (2010) showed that infants learned minimal pairs if the novel words were embedded in sentence frames (i.e. "look at the bih"), or if infants first saw familiar objects (e.g. cat, shoe) labeled. Infants can also succeed with familiar minimal pairs (e.g. ball and doll) (Fennell, 2012). These studies suggest that 14-month-olds can attend to fine phonetic detail when it is evident that the task involves labeling objects.

Adding talker variability during familiarization also helps 14-month-olds learn novel minimal pairs (Galle, Apfelbaum, & McMurray, 2015; Hoehle et al., 2020; Rost & McMurray, 2009). Rost and McMurray (2009) first demonstrated that hearing the novel minimal pair from 18 different talkers (vs. one identical token) allowed infants to successfully map the novel words to distinct objects (see Quam, Knight, and Gerken (2017) for replication in English and Hoehle et al. (2020) in German). Highly variable speech from a single talker too helps infants succeed[1] (Galle et al., 2015). However, there are bounds on the types of variability that help infants succeed. For instance, variability in VOT (Voice Onset Time, which is lexically contrastive in English) was not found to support infant learning of novel minimal pairs (Rost & McMurray, 2010). While many of these

---

[1] While other studies have used multiple tokens from a single speaker (see e.g. Werker et al., 2002; Fennell & Waxman, 2005; Fennell & Waxman, 2012), none of them were intended to maximize variability were typically the same length and produced in the same child-directed-speech contour (e.g. "in an infant-directed, rise-fall intonational phrase", Fennell & Waxman, 2005)

prior studies offer compelling accounts of the mechanisms by which variability influences learning (e.g. Hoehle et al., 2020), taken together, we see these prior studies as suggesting that adding talker variability helps infants attend to the features that remain consistent between- and within-talkers: the distinguishing sounds at the onset of the words

Beyond supporting minimal pair learning, talker variability has also been found to be beneficial for infants learning novel phonotactic patterns (Seidl, Onishi, & Cristia, 2014), integrating input across talkers (Estes & Lew-Williams, 2015), and recognizing words produced by new talkers (Bulgarelli & Bergelson, 2022). However, talker variability can prove challenging for learners. For example, preschoolers and adults are slowed down in recognizing familiar words when the talker changes between trials (Jusczyk, Pisoni, & Mullennix, 1992; Ryalls & Pisoni, 1997), and younger infants often struggle with recognizing words in their native language (e.g. bike, tree) when the talker changes (Houston & Jusczyk, 2000). Similarly, while 8-month-olds can form a robust word-object link when trained with a single token of a word, adding talker variability may lead to overly broad representations about how a word could sound, for example leading infants to accept new words as viable labels for the trained object (Bulgarelli & Bergelson, 2022).

The exact conditions under which talker variability influences learning are still unknown (see Bulgarelli & Weiss, 2021; see also Raviv, Lupyan, & Green, 2022 for the roles of variability across domains). However, an emerging theme is that while talker variability should not be *necessary* for learning, "irrelevant" variability may help identify task-relevant dimensions (Rost & McMurray, 2009; see also Raviv et al., 2022). In the current study, we take this one step further, asking whether "irrelevant" variability influences learning if the "task-relevant" dimensions are already clear. That is, with dissimilar words like "neem" and "lof", talker variability may not be necessary or helpful for learning.

We investigate whether 14-month-olds can learn distinct-sounding novel words with and without talker variability in the Switch task. Since 14-month-olds have been shown to learn such words when presented with identical tokens of a word by a single talker, talker variability may not affect learning. In this case, we would expect participants in all conditions to learn the novel word-object mappings, operationalized as increased looking when the word-object link is broken. However, it is also possible that talker variability may interfere with these early phases of word learning when it isn't necessary to highlight invariant dimensions. In this case, we would expect that infants trained with talker variability may not learn the novel word-object mappings, i.e. failing to increase their looking time when the trained word-object links are broken.

We include two types of talker variability: variability stemming from multiple talkers (Between-Talker) and from a highly variable single talker (Within-Talker). As previous research has found that these types of variability similarly influence learning (14-month-olds; Rost and McMurray (2009); Galle et al. (2015); Tsui, Byers-Heinlein, and Fennell (2019),; 8-month-olds; Bulgarelli and Bergelson (2022)), we do not expect differences between these two conditions. However, as these types of variability provide different acoustic evidence for how words can vary (e.g. mean pitch varies more between talkers than within talkers), both in the real world (Bulgarelli, Mielke, & Bergelson, 2021) and in

the lab (Galle et al., 2015), it is in principle possible that they could impact word learning differently here.

Methods

The preregistration https://osf.io/7jxd9/?view_only=aaec3fb9166744e08dba0a410ce57e7c, (https://osf.io/pjfh6/?view_only=c839cf47ba8e4eba84ff23ea1fdf6a25), as well as stimuli, data, and code used to create this manuscript are posted on OSF: https://osf.io/xnpjk/?view_only=5924cc40a45c433aa36d8ea1d9017996. An a priori power analysis (see preregistration) found that 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 et al., 2019), and is what we use here.

Participants

Our final sample was made up of 54 13–15 month olds (mean age = 14.19 months; 34 female, 20 male). All participants were full term (>37 weeks gestation), monolingual (at least 75% exposure to English by parental report), and had no known hearing or vision difficulties. Participants were recruited from the area surrounding a university in the southeastern United States and using childrenhelpingscience.com. Parents provided consent on behalf of themselves and their infants and were provided with a $5 amazon giftcard as a thank you for participating. 87% of the infants were White or Caucasian, 13% identified as multiracial, multiethnic, other, or preferred not to report. Maternal education ranged from some college to advanced degree (some trade school, professional training, or college: n = 2; associate or bachelor's degree: n = 16; advanced degree: n = 32, n = 4 did not report this information). An additional 22 infants were excluded due to fussiness (n = 10), technical difficulties (n = 5), or not habituating or meeting looking criteria on test trials (n = 7). All procedures involving human subjects were approved by the Institutional Review Board and met the guidelines laid down by the Declaration of Helsinki.

Design

The experiment consisted of the standard, two-word switch task (see Stager and Werker (1997); Werker, Cohen, Lloyd, and Stager (1998)), wherein participants are habituated to two novel word-object pairs. Habituation occurred under one of three talker-variability conditions. In the *No-Talker-Variability* condition, infants heard a single prototypical child-directed token of each novel word produced by a single female talker. In the *Within-Talker-Variability* condition, infants heard 12 highly-variable tokens of each novel word produced by a single female talker. Finally, in the *Between-Talker-Variability* condition, infants heard 10 different female talkers produce each of the novel words. The test phase queried whether infants would notice a break to the word-object link. All infants saw three test trials: *Same*, *Switch* and *Novel*; see Figure 1.

Stimuli

Stimuli consisted of four familiar items (apple, ball, shoe, dog), three novel items (object1 - a kitchen tool, object2 - a dog toy, object3 - interlocked disks), two novel words ('neem' and

'lof'); and an animated attention-getter paired with a jingle. Visual stimuli were videos of the familiar items and novel objects looming on the screen, ranging from 50–90% in height and 30–50% in width of the display.

The auditory stimuli, consisting of recordings of the familiar item labels and the novel words, were identical to those used in Bulgarelli and Bergelson (2022). Each word was recorded by 10 female young adults. These auditory stimuli deliberately maximized acoustic differences stemming from (highly-variable) within- and between- talker variability, and by design varied in multiple dimensions. 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 et al. (2015)); two female talkers did the same for the familiar items. Each token was then spliced and embedded in silence to create 2s long files (see OSF).

One of 2 female talkers (counterbalanced across participants) were used for familiarization in the *No-Talker-Variability* condition and the *Within-Talker-Variability* condition, and for test trials across all conditions. Ten female talkers (including the 2 just mentioned) were used for familiarization in the *Between-Talker-Variability* condition.

Procedure

The experiment was run over Zoom using Habit 2 (Oakes, Sperka, DeBolt, & Cantrell, 2019). After consent, infants sat in their caregiver's lap facing the computer or laptop in their homes. The experimenter shared their screen and ensured that all that was visible on the participants' screen was the experiment (e.g. participants could not see the video of themselves or the experimenter, and the screen was in full screen mode). Parents were asked not to direct their infants' attention and to keep the infant on their lap facing the computer if possible. As the experimental sounds were transmitted through the experimenter's computer[2], the experimenter wore noise cancelling headphones during the study to minimize access to the auditory stimuli (though it was impossible to be completely unaware of the stimuli, see Bulgarelli and Bergelson (2022) for reliability of this methodology and comparison to in-lab testing).

The experiment began with a 9 point calibration to help the experimenter gauge infants' looking patterns to each edge of the screen. Following calibration, there were three types of trials: warm-up trials, habituation trials, and test trials. Each trial began with an attention-getter directing infants' gaze to the monitor. Trials lasted up to 14 seconds, and infants could hear up to 7 instances of the presented word during any trial. During these trials, infants' looking behavior was live-coded by an experimenter who pressed a button while infants looked at the screen. Each trial ended when the infant looked away for more than 2 seconds after looking at the screen for at least 1 second after trial onset.

---

[2]When we first started data collection over Zoom, participants could only hear the sound played by the experimenter if the experimenter's microphone was unmuted, and thus both the experimenter and the participant could hear the stimuli.

Warm-up trial

On each trial (n=4), infants saw one familiar objects (apple, ball, shoe, dog) and heard it labeled. We included this phase as previous research has shown that this highlights the referential nature of the task, i.e. that the objects on the screen are being labeled (Fennell & Waxman, 2010).

Habituation trials

After the warm-up trials, the habituation phase began. Infants viewed trials labeling one object at a time and the corresponding looming video (e.g. some trials labeled object-1, others labeled object-2). Trial order was pseudo-randomized, such that no more than two trials in a row repeated the same word-object pairing[3]. This phase continued until the habituation criteria was reached: when looking time to the last four trials was half as long as looking time to the first four trails, using a sliding window (Casasola & Cohen, 2000); and could last between 5 and 30 trials. All participants met our habituation criteria.

Test trials

The test phase consisted of 2 test trials and a control trial. The *Same* trial matched habituation - e.g. object-1 (kitchen tool) labeled with word-1 (neem). The *Switch* trial presented a mismatch of the word-object pairs from habituation - e.g. object-1 (kitchen tool) now paired with the word-2 label (lof). If infants have learned the word-object link, they should increase their looking time when the link is broken during the *Switch* trial. On the *Control* trial, infants saw a completely novel object, and heard it labeled with a trained word - e.g. object-3 (interlocked disks) and word-1 (neem). The *Control* trial is used to confirm that a lack of increase in looking time to the *Switch* trial is not simply because infants have lost interest in the experiment. Since the object used in the *Control* trial has never been used in the experiment, infants are expected to increase their looking time regardless of condition. The *Control* trial always happened last, but the *Same* and *Switch* trials were counterbalanced across participants. The tokens of the words used at test were always ones heard during the experiment. For half of the participants, the *Same* and *Switch* trial featured object-1 and for the other half the *Same* and *Switch* trials featured object-2.

Data analysis

We used RStudio (RStudio Team, 2019) and R [Version 4.0.2; R Core Team (2020)] to generate this manuscript, along with all figures and analyses; 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* trial and the *Control* trial differed from the *Same* 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 as well as by-Subject random intercepts[4]. Thus, the model formula was as follows:

---

[3]See Supplementals for additional experimental conditions highlight the importance of pseudorandomization.

$$LookingTime \sim TestTrialType \times HabituationCondition + (1|Subj)$$

The *Same* trial always served as our baseline, so the TestTrialType variable was contrast coded to compare the *Switch* test to the *Same* test, and the *Control* test to the *Same* test separately. The HabituationCondition variable used orthogonal contrasts to compare both talker variability conditions to the No Variability condition, and then the *Within-Talker-Variability* and *Between-Talker-Variability* conditions to each other.

Results

Habituation

Participants habituated after an average of 14.80 trials (sd = 5.60). This did not vary by *Talker-Variability* condition, $F(2,51) = 0.36$, $p = .702$.

Test Trials

Results can be found in Figure 2 and Table 1. Neither of the contrasts comparing *Talker-Variability* conditions were significant (*No-Variability* compared to both talker variability conditions together: $t(51) = 1.53$, $p = .131$; *Within-Talker* compared to *Between-Talker*: $t(51) = -0.03$, $p = .976$), suggesting that looking time overall did not vary as a function of habituation condition. We first report results comparing looking time to the *Switch* trial relative to the *Same* trial, the most direct test of whether infants learned the word-object links. There was no overall difference in looking time to the *Switch* trial, $t(102) = -1.43$, $p = .156$ ($M_{Switch} = 6,204.89$ms, $SD_{Switch} = 3,877.35$; $M_{Same} = 5,277.17$ms, $SD_{Same} = 3,005.90$). There was, however, a significant interaction between *Switch* vs. *Same* looking time and whether infants heard talker variability during training, $t(102) = 2.70$, $p = .008$. Participants in the *No-Variability* condition significantly increased their looking time to the *Switch* trial ($M_{Switch} = 8,284.83$, $SD_{Switch} = 3,859.35$) relative to the *Same* trial ($M_{Same} = 5,617.06$, $SD_{Same} = 2,787.13$), $t(30.94) = 2.38$, $p = .024$, while those who heard talker variability during training did not, $t(69.20) = 0.07$, $p = .941$ ($M_{Switch} = 5,164.92$, $SD_{Switch} = 3,491.95$; $M_{Same} = 5,107.22$, $SD_{Same} = 3,133.70$). This suggests that infants trained with talker variability failed to notice when the trained word-object link was broken in the *Switch* trial. *Switch* vs *Same* looking time did not differ for participants in the *Within-Talker* and *Between-Talker* conditions ($t(102) = 0.58$, $p = .563$).

We next report results comparing looking time to the *Control* trial relative to the *Same* trial. Since the *Control* trial uses a completely novel object, this serves as a comparison to ensure that infants do indeed increase their looking time when a brand new object is introduced. As expected and consistent with previous research (e.g. Galle et al., 2015; Rost & McMurray, 2009) participants showed a significant increase in looking time to the *Control* trial relative to the *Same* trial, $t(102) = 5.70$, $p < .001$ ($M_{Same} = 5,277.17$ms, $SD_{Same} = 3,005.90$; $M_{Control} = 8,534.78$ms, $SD_{Control} = 4,226.34$), which did not vary as a function of talker variability

[4]A model including random effects of order and item used at test approached singularity, so we simplified to the random effects structure above.

during training (*No-Talker-Variability* vs. talker variability conditions: $t(102) = -1.63$, $p = .106$; *Within-Talker-Variability* vs. *Between-Talker-Variability*: $t(102) = -0.18$, $p = .861$).

Exploratory analyses

As described in our preregistration, we also considered the role of age, receptive and productive vocabulary, and knowledge of the warm-up words. These variables were linked with task performance; see Supplementals for full results. Infants in our sample were reported to understand 89.78 words (SD = 71.51), and produce 12.02 words (SD = 14.29).

Comparison to previous research

While previous research found that talker variability during training helped infants learn novel minimal pairs (Galle et al., 2015; Hoehle et al., 2020; Rost & McMurray, 2009), our results demonstrate that talker variability during training actually impairs learning of distinct sounding words. We next used publicly available data from the studies reported in Hoehle et al. (2020)[5] to test whether the pattern of results we find with English-learning 14-month-olds taught 2 dissimilar novel words (neem and lof) is statistically different from that of German-learning 14-month-olds taught 2 novel words that are a minimal pair (buk and puk). The methods employed by the present study and Hoehle et al. (2020) are almost identical, allowing direct comparisons of the looking time patterns. Since Hoehle et al. (2020) did not conduct a *Within-Talker-Variability* condition, we only compare data from the *No-Variability* and *Between-Talker-Variability* conditions[6].

We ran a mixed effects model to test whether looking time varied as a function of Trial Type (i.e. whether the *Switch* trial differed from the *Same* trial), habituation condition (*No-Talker-Variability*, *Between-Talker-Variability*), word similarity (dissimilar words (i.e. our data), minimal pairs (i.e. Hohle et al.'s data)), and the interactions between them, alongside Subject-level random intercepts:

$$LookingTime \sim TestTrialType * HabituationCondition * WordSimilarity + (1|Subj)$$

Results are visualized in Figure 3, full model output is in Table 2. Overall, participants marginally increased their looking time to the *Switch* trial (Mean$_{Switch}$ = 7,165.46, SD$_{Switch}$ = 4,327.10; Mean$_{Same}$ = 5,927.81, SD$_{Same}$ = 3,428.28), $t(132) = 1.91$, $p = .058$. There was also a significant interaction between Talker Variability condition and word similarity, $t(132) = -2.41$, $p = .017$, which can best be interpreted in light of a significant three way interaction between test trial, Talker Variability condition and Word Similarity, $t(132) = -2.11$, $p = .037$. Participants in the *No-Talker-Variability* condition only increased their looking time to the *Switch* trial when the words were distinct, $t(30.94) = 2.38$, $p = .024$, but not when they were minimal pairs $t(31.53) = -0.56$, $p = .581$. The opposite pattern was true for participants in the *Talker-Variability* conditions, who increased their looking time to the *Switch* trial when the words were minimal pairs $t(24.34) = 2.38$, $p = .025$, but not when the words were distinct

---

[5]Unfortunately, the raw data from Galle et al., (2015) and Rost & McMurray (2009) are not available to include in this comparison (McMurray, personal communication).
[6]We note that while our between-talker-variability condition included 10 female talkers, Hohle et al., (2020)'s between-talker condition included 18 different male and female talkers.

$t(32.91) = -1.04$, $p = .307$. This suggests that the impact of talker variability on learning in the Switch task is dependent on whether the words to be learned are dissimilar or minimal pairs.

General Discussion

The current study tested whether talker variability influenced 14-month-olds' ability to learn distinct sounding novel words ("lof" and "neem"). While 14-month-olds learned the word-object links when the words were presented using a single identical token, they did not when the word exposure featured talker variability. Thus, while talker variability can help infants learn similar-sounding minimal pairs (Galle et al., 2015; Hoehle et al., 2020; Quam et al., 2017; Rost & McMurray, 2009), it appears to interfere with learning when the words are distinct.

Why is variability sometimes helpful for word learning and sometimes challenging to contend with? Our results suggest that talker variability may be particularly beneficial when it highlights the task-relevant dimension (see Raviv et al., 2022). For example, VOT was the meaningful feature that differed between minimal pairs used in previous studies (e.g. /buk/ and /puk/; Rost and McMurray (2009); Stager and Werker (1997)). Talker variability highlighted what remained consistent across instances and what changed – allowing infants to appropriately attend to what differed (i.e. the word's initial phonemes). In the current study, the words /neem/ and /lof/ vary in many ways, and talker variability may make it harder for infants to zero in on the features they should attend to. Indeed, other work found that variability structured by talker gender did not support minimal pair learning (Quam et al., 2017), possibly because it became harder to determine which variable to attend to. This suggests talker variability may be particularly beneficial when it highlights single feature changes rather than systematic patterns.

A related possibility is that talker variability that does not highlight invariance between two words *increases* task difficulty. This is consistent with the "Resource Limitation Hypothesis" (Stager & Werker, 1997), which proposes that infants have limited resources they can devote to a task, and forgo attention to detail if tasks exceed those resources (e.g. fine phonetic detail when learning minimal pairs). Support for this hypothesis comes from the fact that infants with larger vocabularies are more likely to succeed on the minimal pair task at 14 months (Werker, Fennell, Corcoran, & Stager, 2002), as larger vocabularies suggest better word learning skills that may confer more capacity to devote to the challenging task. Similarly, referential cues that highlight that the task involves labeling objects (Fennell & Waxman, 2010), or allowing infants to play with the novel objects in advance (Fennell, 2005), allows infants to succeed by reducing task demands. As infants age they also gain additional resources, making learning scenarios that were once difficult, easier (e.g. by 17 and 20 months infants succeed on the minimal pair task (Werker et al., 2002))[7].

---

[7]A further possibility is that testing over Zoom rather than in person led to this pattern of results. We find this unlikely given prior working showing convergent patterns for studies conducted in both contexts (see Bulgarelli & Bergelson, 2022; Chuey et al., 2021), and more broadly (Bacon, Weaver, & Saffran, 2021; e.g. Schidelko, Schünemann, Rakoczy, & Proft, 2021).

Fennell and Waxman (2010) also suggests that talker variability was helpful in previous experiments because—like referential labels highlighting the nature of the task—hearing multiple people use the same label for novel objects demonstrates social convergence. And yet, in our study infants failed to learn in the talker variability conditions despite both of these components (warm-up referential label trials and talker variability). Combined with previous research showing that highly-variable speech from a single talker can help infants succeed (Galle et al., 2015), our results suggest that the benefits of talker variability during learning are not always attributable to different talkers providing socially-convergent evidence.

Rather, we propose that introducing "unnecessary" talker variability may have made the task too resource-intensive for infants, making it harder for them to both process talker variability and learn the novel word-object mappings. We suggest that the effect of talker variability may be dependent on a combination of the infant's resource limitations (which depends on age and task difficulty) and the usefulness of the variability itself. Future work could explicitly test this: if a task exceeds the infant's resources, then acoustic variability is expected to be helpful it if can highlight the task-relevant dimension; otherwise, it is predicted to negatively impact performance. While determining resource "quantity" is certainly not trivial, patterns of success and failure in habituating, and tracking labels and referents can serve as proxies. Lastly, on this account, we'd predict that if infant's resources exceed task demands, the addition of talker variability should not matter, regardless of whether it highlights an invariant dimension. Consistent with this possibility, 7.5-month-olds distinguish native language consonant contrasts whether or not they are produced with talker variability (Quam, Clough, Knight, & Gerken, 2020). Future work could test this idea by asking whether talker variability would influence learning of distinct-sounding words for infants >14 months, who have greater linguistic, cognitive, and social resources.

Our study also extended previous research showing that *Within* and *Between* talker variability similarly impact learning (Bulgarelli & Bergelson, 2022; Galle et al., 2015; Rost & McMurray, 2009; Tsui et al., 2019). This is somewhat surprising, as *Within* and *Between* talker variability provide different acoustic information, both in the real world (Bulgarelli et al., 2021) and the lab (see Galle et al., 2015). And yet, we found that both types of talker variability failed to support learning distinct words, while identical single-talker tokens succeeded. Our stimuli were intended to maximize variability within and between talkers, but the exact amount of variability needed to effect learning in this context is unknown. For example, Werker et al. (2002) report using 10 exemplars from one talker, but find that 14-month-olds fail to learn minimal pairs. They specify, though, that these exemplars were recorded in an infant-directed, rise-fall intonation, and that they were each approximately the same length. However, adding highly-variable within-talker tokens influenced learning here, as well as in studies by Bulgarelli and Bergelson (2022) and Galle et al. (2015). Thus, while each instance of a word produced by a single talker is going to vary slightly (see Bulgarelli et al., 2021; Peterson & Barney, 1952), a minimum amount of variability is likely necessary for within-talker variability to exhibit these effects. The current study adds to a growing literature highlighting that while infants are certainly learning words in the first and second year of life, this learning–especially in its earliest phases–is fragile and indeed easily extinguishable by adding a prevalent feature of daily life: talker variability. We look forward

to future research further uncovering the complex interactions between early learning and variability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bacon D, Weaver H, & Saffran J (2021). A Framework for Online Experimenter-Moderated Looking-Time Studies Assessing Infants' Linguistic Knowledge. Frontiers in Psychology, 12. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2021.703839

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

Bergelson E (2020). The Comprehension Boost in Early Word Learning: Older Infants Are Better Learners. Child Development Perspectives, 0(0), 1–8. 10.1111/cdep.12373

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. 10.1073/pnas.1113380109 [PubMed: 22331874]

Bulgarelli F, & Bergelson E (2022). Talker variability shapes early word representations in English-learning 8-month-olds. Infancy, (December 2021), 1–28. 10.1111/infa.12452

Bulgarelli F, Mielke J, & Bergelson E (2021). Quantifying Talker Variability in North-American Infants' Daily Input. Cognitive Science, 46(1), e13075. 10.1111/cogs.13075 [PubMed: 34971003]

Bulgarelli F, & Weiss DJ (2021). Desirable Difficulties in Language Learning? How Talker Variability Impacts Artificial Grammar Learning. Language Learning, 1–37. 10.1111/lang.12464

Casasola M, & Cohen LB (2000). Infants' association of linguistic labels with causal actions. Developmental Psychology, 36(2), 155–168. 10.1037/0012-1649.36.2.155 [PubMed: 10749073]

Chuey A, Asaba M, Bridgers S, Carrillo B, Dietz G, Garcia T, … Gweon H (2021). Moderated Online Data-Collection for Developmental Research: Methods and Replications. Frontiers in Psychology, 12. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2021.734398

Estes KG, & Lew-Williams C (2015). Listening Through Voices: Infant Statistical Word Segmentation Across Multiple Speakers. Developmental Psychology, 51(11), 1–12. 10.1037/a0039725 [PubMed: 25546594]

Fennell CT (2005). Infant attention to phonetic detail in word forms: Knowledge and familiarity effects. Dissertation Abstracts International: Section B: The Sciences and Engineering, 66(2-B), 1196. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2005-99016-014

Fennell CT (2012). Object Familiarity Enhances Infants' Use of Phonetic Detail in Novel Words. Infancy, 17(3), 339–353. 10.1111/j.1532-7078.2011.00080.x [PubMed: 32693488]

Fennell CT, & Waxman SR (2010). What Paradox ? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning. Child Development, 81(5), 1376–1383. 10.1111/j.1467-8624.2010.01479.x [PubMed: 20840228]

Galle ME, Apfelbaum KS, & 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). 10.1080/15475441.2014.895249

Hoehle B, Fritzsche T, Mess 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. 10.1111/desc.12950

Houston DM, & Jusczyk PW (2000). The role of talker-specific information in word segmentation by infants. Journal of Experimental Psychology: Human Perception and Performance, 26(5), 1570–1582. 10.1037/0096-1523.26.5.1570 [PubMed: 11039485]

Jusczyk PW, Pisoni DB, & Mullennix JW (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. Cognition, 43, 253–291. 10.1016/0010-0277(92)90014-9 [PubMed: 1643815]

Oakes LM, Sperka D, DeBolt MC, & Cantrell LM (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. 10.3758/s13428-019-01244-y [PubMed: 31012062]

Pater J, Stager C, & Werker J (2004). The Perceptual Acquisition of Phonological Contrasts. Language, 80(3), 384–402. Retrieved from https://www.jstor.org/stable/4489718

Peterson GE, & Barney HL (1952). Control Methods Used in a Study of the Vowels. The Joual of the Acoustical Society of America, 24(2), 175–184. 10.1121/1.1906875

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

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

R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raviv L, Lupyan G, & Green SC (2022). How variability shapes learning and generalization. Trends in Cognitive Sciences, 1–22. 10.1016/j.tics.2022.03.007

Rost GC, & McMurray B (2009). Speaker variability augments phonological processing in early word learning. Developmental Science, 12(2), 339–349. 10.1111/j.1467-7687.2008.00786.x.Speaker [PubMed: 19143806]

Rost GC, & McMurray B (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. Infancy, 15(6), 608–635. 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 BO, & Pisoni DB (1997). The Effect of Talker Variability on Word Recognition in Preschool Children. Dev Psychol., 33(3), 441–452. 10.1037/0012-1649.33.3.441 [PubMed: 9149923]

Schidelko LP, Schünemann B, Rakoczy H, & Proft M (2021). Online Testing Yields the Same Results as Lab Testing: A Validation Study With the False Belief Task. Frontiers in Psychology, 12. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2021.703238

Seidl A, Onishi KH, & Cristia A (2014). Talker Variation Aids Young Infants' Phonotactic Learning. Language Learning and Development, 10(4), 1–24.

Stager CL, & Werker JF (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. Letters to Nature, 381–383. 10.1038/41102

Tincoff R, & Jusczyk PW (2012). Six-Month-Olds Comprehend Words That Refer to Parts of the Body. Infancy, 17(4), 432–444. 10.1111/j.1532-7078.2011.00084.x [PubMed: 32693484]

Tsui ASM, Byers-Heinlein K, & Fennell CT (2019). Associative word learning in infancy: A meta-analysis of the Switch task. Developmental Psychology, 55(5), 934–950. 10.1037/dev0000699 [PubMed: 30730174]

Werker JF, Cohen LB, Lloyd VL, & Stager CL (1998). Acquisition of Word-Object Associations by 14-Month-Old Infants. 34(6), 1289–1309.

Werker JF, Fennell CT, Corcoran KM, & Stager CL (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. Infancy, 3(1), 1–30. 10.1207/S15327078IN0301_1
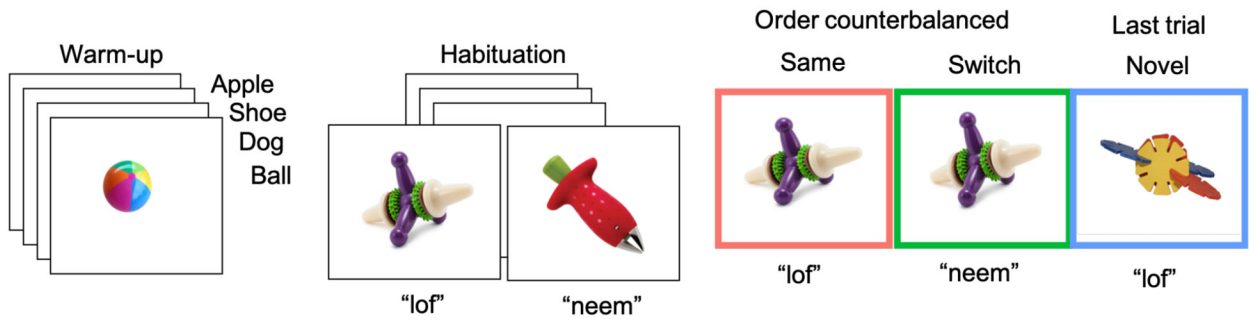
Figure 1.
Example of experimental procedure. Colored boxes correspond to data in subsequent figures.
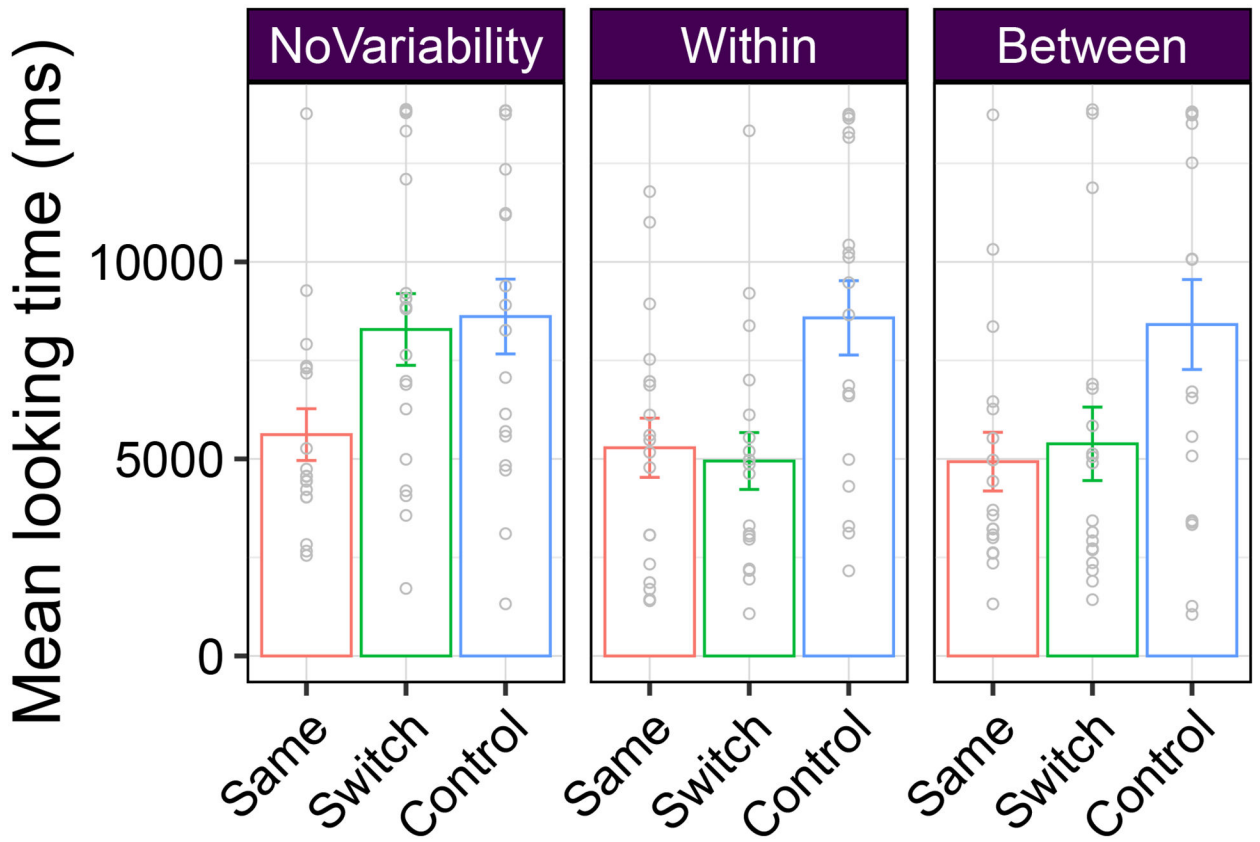
Figure 2.
Experiment results. Bars depict mean looking time (y-axis) across test trials for participants in all three conditions (x-axis). Circles indicate individual data points, error bars reflect standard error. Participants in the No-Talker-Variability condition dishabituated to the Switch and Novel trials. Participants in both Talker Variability conditions only dishabituated to the Novel trial.
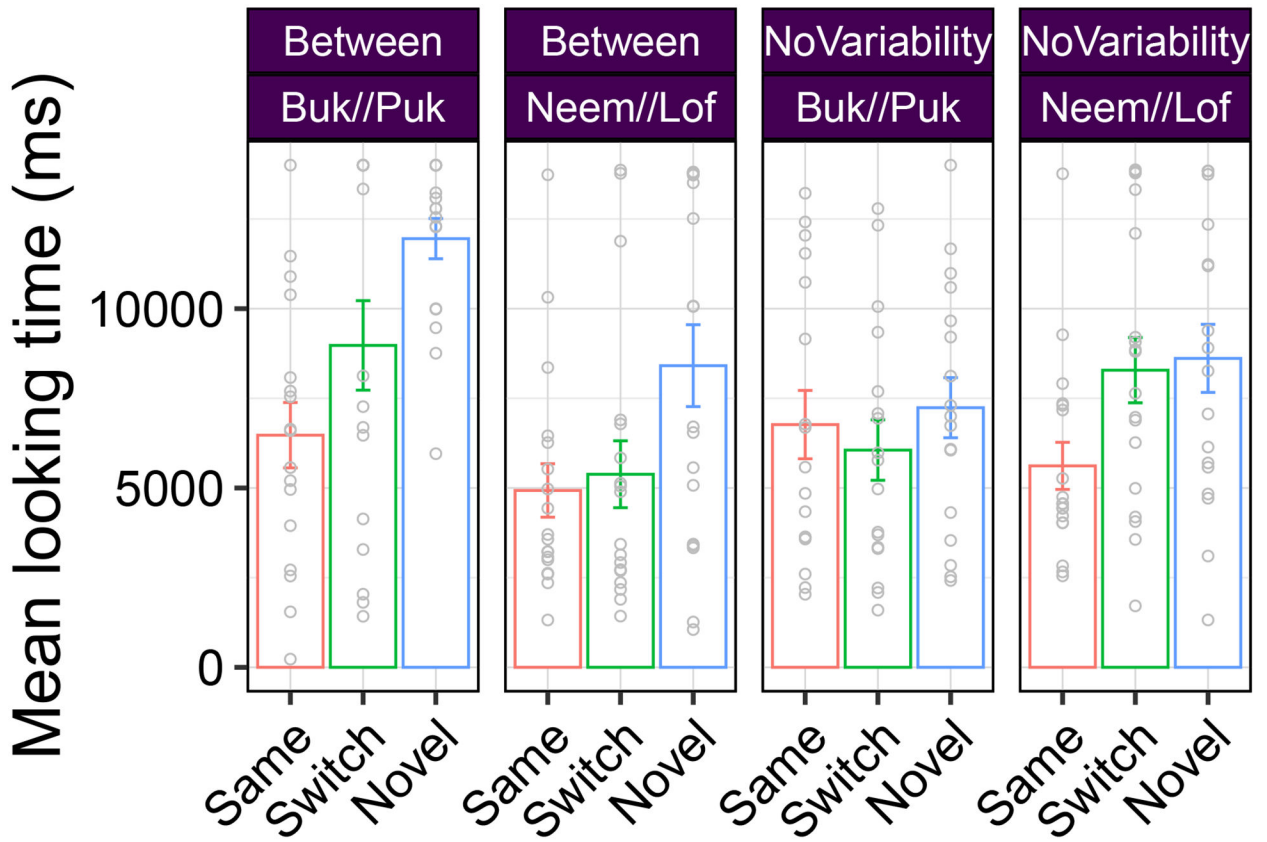
Figure 3.
Comparison of the No-Talker-Variability and Between-Talker conditions in the current experiment (the same data as in Figure 1) and in Hohle et al., (2020). Bars depict mean looking time (y-axis) across test trials for participants in all conditions (x-axis). Circles indicate individual data points, error bars reflect standard error.

Table 1

Fixed effects and Cohen's d for model presented in the results section. X in predictor name indicates interaction. 'SameSwitch' refers to comparison between Same and Switch trial, 'SameControl' refers to comparison between Same and Control trial. 'TalkerVariability' refers to comparison between No Talker Variability and Talker Variability conditions, 'WithinBetween' compares the Within talker condition to the Between talker condition. SE is pooled for each predictor

term	estimate	std.error	statistic	p.value	d
(Intercept)	6,672.28	383.88	17.38	<.001	NA
SameSwitch	−467.39	326.67	−1.43	0.156	−0.283
SameControl	1,862.50	326.67	5.70	<.001	1.129
TalkerVariability	1,249.31	814.32	1.53	0.131	0.430
WithinBetween	−14.16	470.15	−0.03	0.976	−0.008
SameSwitch:TalkerVariability	1,870.61	692.98	2.70	0.008	0.535
SameControl:TalkerVariability	−1,131.14	692.98	−1.63	0.106	−0.323
SameSwitch:WithinBetween	232.35	400.09	0.58	0.563	0.115
SameControl:WithinBetween	−70.34	400.09	−0.18	0.861	−0.035

Table 2

Fixed effects model table for comparison of results in the current manuscript and those reported by Hohle et al., (2020) using minimal pairs. 'SameSwitch' compares Same to Switch trial, 'TalkerVariability' compares No Talker Variability to Between Talker Variability conditions, 'WordSimilarity' refers to minimal pairs vs. dissiminar words. X in the variable name refers to an interaction.

term	b	95% CI	t	df	p
Intercept	6,561.12	[5,924.94, 7,197.31]	20.40	132	< .001
SameSwitch	614.08	[−22.10, 1,250.27]	1.91	132	.058
TalkerVariability	−120.54	[−756.73, 515.65]	−0.37	132	.708
WordSimilarity	−507.11	[−1,143.30, 129.08]	−1.58	132	.117
SameSwitchXTalkerVariability	124.68	[−511.50, 760.87]	0.39	132	.699
SameSwitchXWordSimilarity	165.88	[−470.31, 802.06]	0.52	132	.607
TalkerVariabilityXWordSimilarity	−776.39	[−1,412.58, −140.21]	−2.41	132	.017
SameSwitchXTalkerVariabilityXWordSimilarity	−678.61	[−1,314.80, −42.43]	−2.11	132	.037