Title

Minimal gains for minimal pairs: Difficulty in learning similar-sounding words continues into preschool

Authors

Creel, Sarah C

Frye, Conor I

Publication Date

2024-04-01

DOI

10.1016/j.jecp.2023.105831

Journal of Experimental Child Psychology 240 (2024) 105831

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Minimal gains for minimal pairs: Difficulty in learning similar-sounding words continues into preschool

Sarah C. Creel *, Conor I. Frye

Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093, USA

A R T I C L E I N F O

Article history:
Received 9 June 2023
Revised 21 November 2023
Accepted 27 November 2023

Keywords:
Phonological acquisition
Minimal pairs
Eye tracking
Word learning

A B S T R A C T

A critical indicator of spoken language knowledge is the ability to discern the finest possible distinctions that exist between words in a language—minimal pairs, for example, the distinction between the novel words *beesh* and *peesh*. Infants differentiate similar-sounding novel labels like “bih” and “dih” by 17 months of age or earlier in the context of word learning. Adult word learners readily distinguish similar-sounding words. What is unclear is the shape of learning between infancy and adulthood: Is there a nonlinear increase early in development, or is there protracted improvement as experience with spoken language amasses? Three experiments tested monolingual English-speaking children aged 3 to 6 years and young adults. Children underperformed when learning minimal-pair words compared with adults (Experiment 1), compared with learning dissimilar words even when speech materials were optimized for young children (Experiment 2), and when the number of word instances during learning was quadrupled (Experiment 3). Nonetheless, the youngest group readily recognized familiar minimal pairs (Experiment 3). Results are consistent with a lengthy trajectory for detailed sound pattern learning in one’s native language(s), although other interpretations are possible. Suggestions for research on developmental trajectories across various age ranges are made.

* Corresponding author.
E-mail address: screel@ucsd.edu (S.C. Creel).

https://doi.org/10.1016/j.jecp.2023.105831
0022-0965/© 2023 The Author(s). Published by Elsevier Inc.

Introduction

A critical indicator of spoken language knowledge is the ability to discern the finest possible distinctions that exist in the language, such as the distinction between the novel words *beesh* and *peesh*. If learners can treat two very similar-sounding words as meaning different things, then certainly they can learn that any other less similar-sounding word pair mean different things. Evidence from infants suggests that such words can be distinguished in lab settings fairly early, and evidence from adults suggests excellent (well above chance) learning of such pairs. However, given dissimilarities between test measures for infant and adult age groups, it is difficult to know how sound distinctions may sharpen over development. Do they stabilize fairly early in childhood, permitting the massive vocabulary growth evident in young children? Instead, do infant findings reflect an early point on a long learning trajectory that may last into adulthood?

These learning trajectory questions are mirrored in other areas of development, and in some cases researchers have uncovered substantial nonlinearities, that is, fairly rapid transitions from low ability to high ability. For example, data from newly walking children suggest that their walking behavior improves nonlinearly, with the biggest improvements in multiple motor parameters occurring during the first few months after walking onset (e.g., Adolph et al., 2003; Bril & Ledebt, 1998). Adolph et al. (2003) wrote that "at certain periods in [walking] development infants 'get a bigger bang for their buck' than at other periods in development" (p. 492). Within language development, there is evidence for a nonlinear "word spurt" in language production (e.g., Goldfield & Reznick, 1990; McMurray, 2007).

The developmental time course of speech perceptual development as it relates to word learning is less clear, although there are some theoretical accounts (discussed below). Multiple studies have tested infants' learning of novel minimal-pair words, such as *bih* and *dih*. Many such studies use the "switch" procedure, where infants are habituated to either one or two label–shape pairs and then a mislabeling (vs. correct labeling) is tested (Werker et al., 2002). For example, Stager and Werker (1997) habituated 14-month-olds to the novel word *bih* plus a picture of Object A and the novel word *dih* plus a picture of Object B, and then they tested whether the infants dishabituated to a mismatched trial (such as *bih* plus Object B) as compared with a matched trial (such as *bih* plus Object A). Longer looking times to a "mislabeling" event are interpreted as learning word–object associations. In this procedure, at 14 months of age, infants can learn dissimilar words like *lif* and *neem* (Stager & Werker, 1997, Experiment 3) but cannot learn similar-sounding words (Stager & Werker, 1997, Experiments 1 and 2), whereas 17- and 18-month-olds can learn similar-sounding words (Werker et al., 2002). That is, as a group, 17-month-olds look significantly longer to a "mislabeled" picture (a *bih* shape while hearing "dih") than to a correctly labeled picture (a *dih* shape while hearing "dih"). Learning may surface even earlier, at 14 months, under propitious conditions (Fennell & Waxman, 2010; Rost & McMurray, 2009, 2010). This result has been replicated across a large number of phonetic features, including consonant place of articulation (Pater et al., 2004; Stager & Werker, 1997), consonant voicing (Rost & McMurray, 2009, 2010), vowel quality (Curtin et al., 2009), and stress (Curtin, 2010), and in both monolingual and bilingual populations (e.g., Fennell & Byers-Heinlein, 2014).

Interestingly, a recent meta-analysis of the switch procedure in children aged 12 to 20 months found smaller effect sizes for studies testing similar-sounding word learning relative to dissimilar word learning (Tsui et al., 2019). This suggests that although similar word learning can take place during the second year of life, it is more challenging than learning less similar words. However, it is not clear what this prefigures for later processing of language sounds, with possibilities ranging from fairly rapid early development sometime after the middle of the second year to gradual, incremental perceptual learning of a language's sound contrasts. Although the experimental evidence is equivocal, some theoretical accounts favor a rapid change explanation. A major framework for understanding differences across tasks and ages for speech processing during infancy is PRIMIR (a developmental framework for Processing Rich Information from Multidimensional Interactive Representations; Werker & Curtin, 2005). PRIMIR posits three filters (innate biases, children's development, and task requirements) that influence children's attention to different planes of organization—perceptual, word form, and phonemes. Werker and Curtin (2005) posited the existence of abstract phonemes and predicted that "once the phoneme plane is firmly established, phonemes should be readily utilized across a variety of tasks. We have argued that this is why the older child is more successful at accessing phonetic detail when learning novel words" (p. 225). Here, "older child" refers to children over 17 months of age as opposed to around 14 months. Although these authors did not use the term "nonlinear," their description sounds like a nonlinear change—before the phoneme plane emerges versus after the phoneme plane emerges, presumably partway through the second year of life.

Other research and theoretical accounts are more consistent with gradual learning. Research on familiar-word recognition among children aged 6 years or older suggests increasing accuracy and sensitivity through 12 years (Hazan & Barrett, 2000) or even 18 years (McMurray et al., 2018). Keep in mind that a familiar word to a 6-year-old is even more familiar (in the sense of more exposures) to a 12-year-old and even more so to an 18-year-old. Munson and colleagues (e.g., Munson, Kurtz, & Windsor, 2005; Munson, Swenson, & Manthei, 2005) presented related findings in speech production. Their findings were based on typically developing as well as clinical populations (developmental language disorder and phonological processing disorder). The argument of Munson, Kurtz, et al. (2005) and Munson, Swenson, et al. (2005) was that vocabulary size potentiates more accurate production because increases in vocabulary contribute to ongoing refinement of phonological categories. This set of findings is more consistent with gradual learning of sound contrasts than with rapid learning during the first few years of life.

Although little work has tested children past infancy on similar word learning, one study with 3- to 5-year-olds found above-chance learning of similar-sounding words like *mard* and *marv* (~70% pointing accuracy where chance = 50%) with an additional talker identity cue; that is, a male speaker always said *mard* and a female speaker always said *marv* (Creel, 2014). However, without comparison with any other benchmark, such as adult performance or dissimilar word learning, it is not clear what this data point represents, particularly given that children in the Creel (2014) study also had talker differences as a cue to word identity.

Difficulty may extend even further than mid-childhood. On the one hand, by adulthood speakers of a language have learned vast numbers of minimal pairs. On the other hand, even adults learning new words sometimes find similar-sounding newly learned words to be more difficult to tell apart than dissimilar-sounding ones (Creel et al., 2006; Creel & Dahan, 2010; Escudero et al., 2016; Magnuson et al., 2003; Pająk et al., 2016; White et al., 2013). Still, adult participants in all the cited studies performed well above chance. Furthermore, these findings must be viewed in the context of task difficulty. Specifically, whereas infants typically learn one or two word–picture associations before testing of similar-sounding words, adults are typically asked to learn large numbers of similar-sounding words at a time (40: Creel et al., 2006; 32: Creel & Dahan, 2010; 8: Escudero et al., 2016; 16: Magnuson et al., 2003; 16: Pająk et al., 2016; 48 [although only 24 per day]: White et al., 2013). Adults might perform near ceiling on infant-like test conditions, that is, if they learned a single word pair at a time (see Experiment 1 below and General Discussion).

In view of evidence of ongoing difficulty during childhood in learning subtle auditory distinctions, Creel and Quam (2015; see also Creel, 2018) posited a *protracted perceptual learning hypothesis.* On this hypothesis, which drew from exemplar approaches to spoken language learning (e.g., Goldinger, 1996, 1998; Johnson, 1997; Munson, Kurtz, et al., 2005; Munson, Swenson, et al., 2005; see also Werker & Curtin, 2005), children begin learning language by storing many traces of individual word forms over long time periods. Only after lengthy learning do regularities that facilitate new word acquisition begin to emerge. This means that brief learning experience might not yield stable representations and that knowledge of a speech sound in one word might not generalize to another word with the same speech sound. This fits with evidence that young children's perceptual skills undershoot those of adults in speech (Creel, 2022; Hazan & Barrett, 2000; McMurray et al., 2018; but see, e.g., Werker & Tees, 1984, for evidence of better infant perception of unfamiliar language sounds) and other domains, including face processing (Anzures et al., 2009; Germine et al., 2011; but see evidence for younger infants' better discrimination of other-race faces [Kelly et al., 2007] and even other-species faces [Pascalis et al., 2002]), accent differentiation (Creel, 2018; Floccia et al., 2009; Girard et al., 2008), and voice learning (Creel & Jiménez, 2012; Fecher & Johnson, 2018).

One piece of information that would better delineate the age trajectory of speech sound learning is how learners at age ranges *between* infancy and adulthood perform at learning similar-sounding words. There is little relevant evidence due to a variety of factors. First is a lack of testing of these questions in populations between infancy and adulthood. A second factor, described in more detail below, is differences in paradigms across age groups (Creel & Quam, 2015). Studies most commonly either (a) use tests that allow detection of a change to an immediately preceding item or series of items (habituation or conditioned head-turn in infants; same–different or similarity ratings in young children) or (b) use familiar words, and both of these might not fully reflect capacity to *learn* similar-sounding words, as described below.

Discrimination tests (habituation, conditioned head-turn, and same–different judgments) might not reflect ability to learn similar-sounding words. These tests of short-term representations of speech sound differences include innumerable infant speech sound habituation studies as well as word similarity judgment studies in older children (e.g., Creel, 2022, with 3- to 5-year-olds; Gerken et al., 1995, with 4-year-olds; Treiman et al., 1998, with 5-year-olds). For habituation paradigms (ba . . . ba . . . ba . . . ba . . . ba . . . *pa), repetition of the standard stimulus may build up short-term representations that are stronger than children's current long-term word knowledge. In a word similarity judgment task where two words are heard back to back, child listeners only need to retain the representation of the first word briefly, long enough to hear the second word for comparison. Thus, such studies may permit better perceptual resolution than tests that tap longer-term memory. For example, if comparing "deev" with "teev" (Creel, 2022), listeners only need to maintain a representation of the first sound ("deev") until they hear the second sound ("teev"), typically no more than a few seconds. However, tasks that require remembering a word–object association ("teev" is the green spiky object) typically require retention over a longer time span. Of course, this might not be true of the switch procedure that is widely used with infants (Stager & Werker, 1997; Werker et al., 2002); it assesses a dishabituation response to a novel picture–word combination (marked with *) that immediately follows a familiarized picture–word combination (. . . bih/A, dih/B, dih/B, bih/A, *bih/B, dih/B), which in principle might be computed in immediate memory (we address this question for our current paradigm in Appendix A).

A different interpretational issue is present for studies of familiar-word recognition. Familiar-word recognition is, on the one hand, highly ecologically relevant. However, it cannot cleanly assess age differences in word encoding ability because older listeners have necessarily had more experience with the familiar words than younger listeners. That is, an adult has heard "doll" and "ball" many more times than a 10-year-old, who has heard those words many more times than a 4-year-old, and so on. Consistent with word familiarity advantages, 14-month-old infants appear to be too young to distinguish novel-word minimal pairs (Stager & Werker, 1997), but they readily distinguish familiar-word minimal pairs (Fennell & Werker, 2003; see related work by Swingley & Aslin, 2002; see Barton, 1976, and Garnica, 1971, for similar evidence from 2-year-olds). Thus, better or more fine-grained familiar-word recognition with increasing age (Hazan & Barrett, 2000; McMurray et al., 2018), although consistent with protracted perceptual learning, might reflect a larger number of learning instances rather than capacity to learn.

Questions of minimal-pair learning always raise the issue of what the exact minimal pairs are: Are single-feature differences generally created equal, or are some "more minimal" (harder to process) than others? A number of researchers have argued during recent years that consonants and vowels play different roles in word learning and recognition. Specifically, consonants are weighted more strongly than vowels (Bonatti et al., 2005; Nazzi, 2005; Nazzi et al., 2016; Nespor et al., 2003) from late infancy onward (but see Højen & Nazzi, 2016, on learning in Danish, a language with an unusual diversity of vowels). Given that much of this research uses French- or Italian-based novel words, and uses either infants or adults, extensions to additional languages and ages are needed. In particular, only a few studies have tested children past infancy (Havy et al., 2011, 2014; Nazzi et al., 2009). Only Nazzi et al. (2009) tested English-learning children near our age range (2.5 years), and that study showed a consonant bias, although English-learning children aged 1.3 to 1.9 years do not show a consonant bias (Floccia et al., 2014; see also Mani & Plunkett, 2007; Swingley, 2016). Thus, we included both consonant and vowel minimal pairs in our stimuli.

The current work

The current set of experiments aimed to differentiate between rapid early improvement and protracted perceptual learning of speech sound properties. We did so by assessing the difficulty of encoding words' sound patterns after infancy by testing a preschool-aged group of children learning minimal-pair words. The range of children tested in the first two experiments was from just under 3 years to over 6 years, representing a wide age span that allowed us to test the influence of age as a continuous factor. Note that previous researchers (Stager & Werker, 1997; Werker et al., 2002) have reported changes in minimal-pair word learning in the switch paradigm from 14 to 17 months, a smaller increase in age (∼20%) than the increase over the span from 3 to 6 years (or even from 3 to 5 years), around a doubling in age. Thus, we expect that our age range should be broad enough to detect improvements with age. In Experiment 1, the goal was to verify that children could learn minimal-pair words. Experiment 2 replicated Experiment 1 but with a distinct word control condition and more child-friendly speech. Finally, Experiment 3 asked whether learning in the youngest children was improved by quadrupling the total number of word tokens heard, putting it on the order of total word tokens heard in previous studies with infants.

In all cases, the prediction of the early asymptote account is fairly clear: If 17-month-olds, or even 14-month-olds, can perform above chance at learning minimal-pair words as referring to distinct pictures, then presumably much older children who are on average triple their age (17 months * 3 = 51 months [∼4.25 years]) will perform with relatively high accuracy. However, if the protracted perceptual learning hypothesis holds, then learning of minimal-pair words should still be relatively weak during the preschool years (and likely beyond), possibly with age-related improvements.

A secondary question addressed here was whether segment type—consonants versus vowels—influences children's ease in minimal-pair learning. As reviewed briefly above, previous work has suggested that English-learning children, unlike learners of other languages (French and Italian), might not show better processing of consonants versus vowels. Because we were interested in the effects of this variable on children's learning, and some researchers (including us) have previously found differences in English-speaking *adults'* word learning as a function of consonant versus vowel overlap (Creel et al., 2006; van Ooijen, 1996), we included both consonant and vowel minimal pairs.

Experiment 1

Here, we asked how well preschool-aged children could learn minimal-pair words. Each child learned two words composing a consonant minimal pair and two other words composing a vowel minimal pair as labels for pictured cartoon referents (see Fig. 1 in Method). At test, children were asked to point to the named cartoon as their eye movements to the cartoons were tracked. Given reported successes of 17-month-olds in learning minimal pairs (e.g., Werker et al., 2002, and many following articles), we initially expected that our much older participants would find this task fairly simple.

Method

Participants

Target sample size was set a priori at $N = 32$. In a previous multi-experiment study using similar-sounding words, Cohen's d for pointing accuracy ranged from 0.70 to 1.63. Selecting a value just below this range, $d = 0.60$, suggests a sample size of 24 to detect bidirectional difference from chance at 80% power (Faul et al., 2007). Increasing this to 32 participants allowed for counterbalancing across word pairs (see below). As a check that good learning was possible given our set of stimuli, we tested a comparison sample of adult monolingual English speakers ($N = 32$) recruited from the university's human participant pool.

For all experiments, children were recruited from a variety of local preschools and day-care facilities. We did not request that caregivers provide ethnicity information. The only restrictions were that the children needed to be native monolingual speakers of English (i.e., parents reported no exposure to or understanding of other languages) and did not have uncorrected visual or hearing impairments.

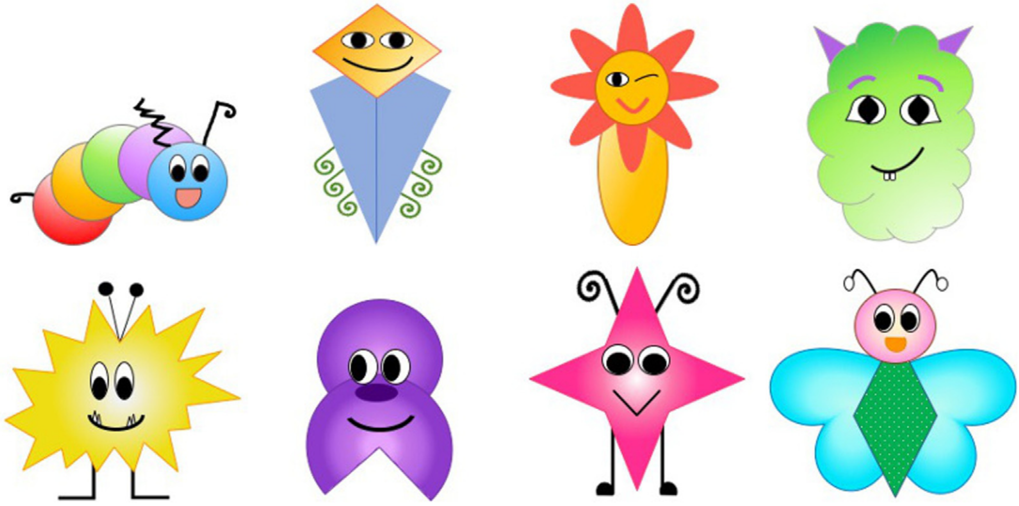

Fig. 1. Cartoon referents used throughout. The cartoons in the top row were used in Experiment 1, all cartoons were used in Experiment 2, and the third cartoons in the two rows were used in Experiment 3.

Note that data were collected prior to the COVID-19 pandemic, meaning that at the time we did not take additional precautionary health measures.

We tested 32 children (19 girls), aged 3 to 6 years (M = 4.72 years, SD = 1.00, range = 3.17–6.86; only 3 children were over 6 years of age). An additional 9 children were tested but replaced due to computer error (n = 3), erroneous sound file presentation (n = 2), experimenter error (n = 1), taking off headphones (n = 1), or loss of interest leading to ending experiment early (n = 2).

This research was approved by our institution's institutional review board.

Stimuli

We used a set of 32 novel consonant–vowel–consonant words (Table 1). Each word differed in a single speech sound from two other words in the set; one had an initial consonant that differed in voicing, and the other had a vowel that differed in quality (tense or lax). This generated 16 consonant minimal pairs and 16 vowel minimal pairs. Across children, all pairs were used as learning stimuli. Isolated tokens of words were recorded by the second author, a young adult male native speaker of English from the western United States. Because stimuli were first used in a longer series of studies with adult listeners, recordings were in adult-directed speech. Recordings were edited and normalized to 70 dB SPL (sound pressure level) in Praat (Boersma & Weenink, 2014). Picture stimuli (Fig. 1, top row) were four colorful cartoon characters that were unfamiliar to children prior to the experiment.

Procedure

Children were tested in a quiet area in their preschool or day-care facility with two experimenters present. One experimenter ran the computer controlling the EyeLink 1000 remote eye tracker (SR Research, Mississauga, Ontario, Canada), and the other ran the experimental presentation computer and entered the children's pointing responses. Experimental presentation was done in Psychtoolbox-3 (Brainard, 1997; Pelli, 1997) using the EyeLink toolkit (Cornelissen et al., 2002). Children wore child-sized KidzGear headphones (www.gearforkidz.com) adjusted to a comfortable listening level. Adults were tested in a quiet room on campus in the principal investigator's lab, also wearing high-quality headphones, after completing another brief experiment designed for children that tested recognition of adult versus child voices.

Children took part in two blocks, each of which consisted of a learning phase with a single word pair followed by a test phase with that word pair (depicted in Table 2). Order of learning the consonant versus vowel minimal pair was counterbalanced across children, as was the assignment of cartoon charac-

Table 1
Experiment 1 novel-word stimuli

vosh /vaʃ/	fosh /faʃ/	beesh /biʃ/	peesh /piʃ/
vush /vʌʃ/	fush /fʌʃ/	bish /bɪʃ/	pish /pɪʃ/
vayfe /veɪf/	fayfe /feɪf/	boove /buv/	poove /puv/
vehf /vɛf/	fehf /fɛf/	buhv /bʊv/	puhv /pʊv/
zodge /zadʒ/	sodge /sadʒ/	dayge /deɪdʒ/	tayge /teɪdʒ/
zudge /zʌdʒ/	sudge /sʌdʒ/	dedge /dɛdʒ/	tedge /tɛdʒ/
zoof /zuf/	soof /suf/	deev /div/	teev /tiv/
zuhf /zʊf/	suhf /sʊf/	dihv /dɪv/	tihv /tɪv/

Note. Consonant voicing minimal pairs are to the left/right, and vowel minimal pairs are above/below. The International Phonetic Alphabet appears within "/ /" for clarity.

ters to labels. In the learning phase of each block, children met two cartoon characters, and each was labeled with one of two words in a minimal pair. On each of 16 randomly ordered learning trials (8 per character), animation sequences programmed in MATLAB depicted one character moving onto the screen and then pausing at the center. It was labeled twice with a 1-s delay between namings (e.g., "Deev. Deev."), and then the creature moved off-screen. Thus, each name was heard 16 times in the learning phase. A single token of each word was used within and across trials. For interest, cartoons' movements (e.g., smoothly moving or "hopping") and screen side (left or right) were varied across trials. Next, in the test phase (16 randomly ordered trials; see Creel, 2014, for a similarly high number of test trials), pictures of the two characters appeared on-screen (left–right location counterbalanced across trials), and one name was spoken. During testing, we recorded children's eye movements and pointing responses to pictures. We collected both measures in an effort to connect related word-learning studies in infants versus older children and adults, which often use disparate measures (e.g., looking measures but not pointing during infancy, pointing accuracy or mouse clicks to pictures in older children and adults). By assessing both simultaneously, we could evaluate how the measures relate to each other and perhaps "translate" between those earlier studies (a topic we took on more fully in another manuscript from our lab that is currently under review; briefly, the correlation across 900 participants between looking proportion and pointing accuracy is ~.70). An experimenter recorded children's pointing response on each trial by clicking the mouse on whichever picture the children pointed to.

Results

Pointing accuracy

To assess pointing accuracy (Fig. 2), we computed a logistic mixed-effects regression in R (R Core Team, 2016) using the *lme4* package (Bates et al., 2015). Logistic regression is appropriate for pointing accuracy data, which are binomially distributed, and mixed-effects models allow both participants and items random effects to be accounted for in the same model. Before analysis, we centered the predictor segment type (consonant minimal pair or vowel minimal pair) and scaled and centered the predictor age, so that the intercept term reflected the grand mean and the effect of segment type reflected the difference between consonant and vowel minimal pairs. Centering variables allows interpretation of effects in the same way as in an analysis of variance (ANOVA). We included maximal random effects: random intercepts and segment type slopes for both participants and words.

The intercept term was significant (B = 0.50, SE = 0.15, z = 3.33, p = .0009), indicating that overall children were performing modestly, but statistically significantly, above chance (M = .60, SD = .16). However, the effect of segment type was not significant (B = 0.14, SE = 0.15, z = 0.93, p = .35), indicating that consonant and vowel minimal pairs were of roughly equal difficulty. Individually, consonant minimal pairs did not exceed chance performance (M = .56, SD = .23, B = 0.35, SE = 0.24, z = 1.44, p = .15),[1]

[1] This number becomes significant (M = .60, SD = .20, p = .02) if we eliminate the 2 participants who appeared to reverse responses (accuracy = .0625 for both). We attribute this response reversal pattern to a strategy of short-term but not long-term storage of word forms, visual forms, and their associations. The logic is that these children likely forgot the word–picture mapping but at test guessed the mapping (wrongly) and largely maintained it throughout the test trials, scoring so far below chance (p = .0005 in a binomial test) that they must have been able to differentiate the two words on consecutive trials.

Table 2
Experiment 1 procedure

Block	Phase	Set 1	Set 2	Set 3	Set 4
1	Learn (16 trials)	*sodge (pic2)* *zodge (pic1)*	*sudge (pic4)* *zudge (pic3)*	zodge (pic1) zudge (pic2)	zoof (pic3) zoohf (pic4)
	Test (16 trials)	*sodge (pic2)* *zodge (pic1)*	*sudge (pic4)* *zudge (pic3)*	zodge (pic1) zudge (pic2)	zoof (pic3) zoohf (pic4)
2	Learn (16 trials)	tedge (pic3) tayge (pic4)	tiv (pic1) teev (pic2)	*tedge (pic4)* *dedge (pic3)*	*tayge (pic2)* *dayge (pic1)*
	Test (16 trials)	tedge (pic3) tayge (pic4)	tiv (pic1) teev (pic2)	*tedge (pic4),* *dedge (pic3)*	*tayge (pic2)* *dayge (pic1)*

Note. Learning took place in two blocks, and each block contained learning trials (16) and testing trials (16). Four (of 16) example stimulus sets are depicted. Pic = picture associated with the label. Italicized cells are different-consonant blocks, and roman (plain) font cells are different-vowel blocks.

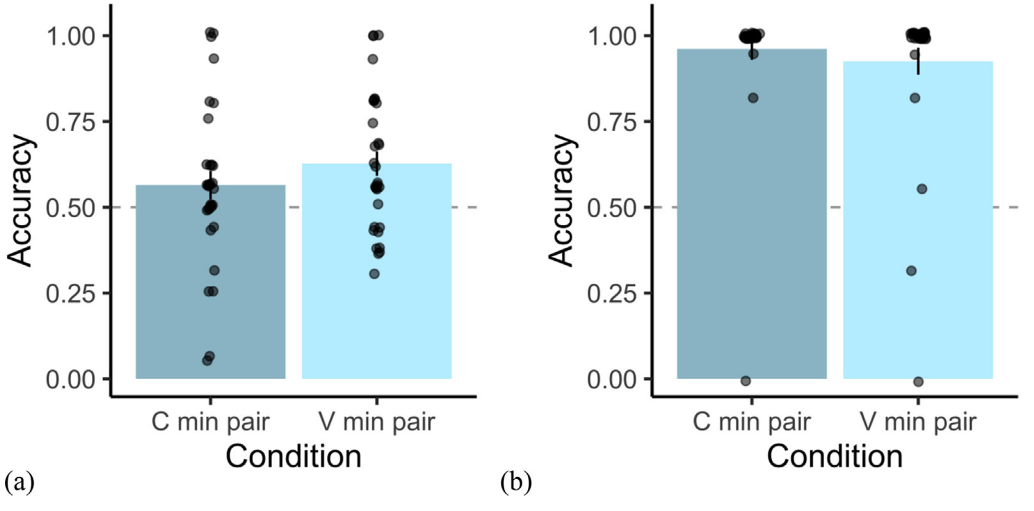

Fig. 2. Experiment 1: Children's pointing accuracy (A) and adults' mouse-clicking accuracy (B), with standard errors. Points are individual participants. Data figures were created using the *ggplot2* package for R (Wickham, 2009). C, consonant; V, vowel; min, minimal.

whereas vowel minimal pairs did (M = .63, SD = .20, B = 0.64, SE = 0.20, z = 3.20, p = .001). Age had a marginal effect (B = 0.27, SE = 0.15, z = 1.85, p = .06), suggesting slightly better performance by older children. Age did not interact with segment type (B = 0.03, SE = 0.14, z = 0.19, p = .85).

Visual fixations: Data processing

The following data processing steps were taken across all experiments. Two details of raw eye tracking data were updated as follows. First, for periods where no eye data were available, we filled in a look to the last known looking position. We took this step because children's pointing movements sometimes obscured the eye tracker camera, and the most parsimonious assumption is that they were continuing to look at whatever they fixated prior to pointing. (This interpolation procedure applied to child looking data only; adult participants responded by clicking a mouse, which did not obscure the camera.) Interpolation of other data types, such as in pupillometry, is not uncommon (see, e.g., Tamási et al., 2017, supplementary material). This is important because counting these periods as looks to nothing may underestimate children's level of recognition, especially if greater certainty leads to earlier pointing. Note that this approach does *not* conflate pointing data and looking data because no information about pointing responses is used. Second, for (rare) trials where children responded especially quickly, we extended the duration of the trial to the full extent of the analysis window (2000 ms). That is, if gaze data ended earlier than 2000 ms post-word onset, the final gaze location was filled in for additional samples to artificially lengthen the trial to 2000 ms. Next, we removed test trials on which children were fixating the monitor less than 50% of the time, which would indicate possible inattention. In the current experiment, this resulted in the removal of 10.7% of trials. Finally, we computed *target advantage* (looks to target minus looks to other picture) in the time window of 200 to 2000 ms, similar to time windows used in Yoshida et al. (2009) and other infant visual world studies (e.g., Swingley & Aslin, 2000, 2002). These scores were aggregated across participants or items to achieve approximately normal distributions. Although some researchers favor analysis at the single-trial level, note that single-trial fixation proportions are neither normal (Gaussian) nor binomial, making it unclear whether linear *or* logistic mixed-effects models are suited to characterizing such data. Although linear mixed-effects models would be more parallel to the logistic mixed-effects models of pointing accuracy, unfortunately they do not tolerate the small number of data points resulting from computing participant or item averages. Thus, we used ANOVAs and *t* tests to compare looks across conditions and with chance.

Visual fixations

To analyze eye tracking data (Fig. 3), we computed participants and items (words) ANOVAs with segment type (consonant minimal pair or vowel minimal pair) as a within-participants and within-items predictor. Only the by-participants model also included age (scaled and centered) as a between-participants predictor; because items were distributed across participants, they did not have markedly different "ages." Segment type was not significant in either participants or items analyses, $F1(1, 30) = 0.10$, $p = .76$; $F2(1, 31) = 0.00$, $p = .99$. Age reached significance, $F1(1, 30) = 4.63$, $p = .04$, suggesting greater looking proportions in older children. The interaction did not approach significance, $F1(1, 30) = 0.02$, $p = .888$).

To examine whether children showed evidence of recognition in their visual fixations, we collapsed analyses over segment type and age and tested fixations versus chance. Target advantage ($M = .077 \pm .154$) exceeded chance, $t1(31) = 2.81$, $p = .009$; $t2(31) = 2.84$, $p = .008$, mirroring the pointing accuracy data. Individually, visual fixations to vowel minimal pairs exceeded chance ($.085 \pm .180$), $t1(31) = 2.67$, $p = .01$; $t2(31) = 2.19$, $p = .04$, whereas visual fixations to consonant minimal pairs did not ($.068 \pm .247$), $t1(31) = 1.56$, $p = .13$; $t2(31) = 1.71$, $p = .10$).[2] Consistent with pointing data, this provides little evidence for consonant superiority in word learning.

Comparison adult learners

These learners were highly accurate ($M = .94$, $SD = .14$),[3] with no significant differences between consonants and vowels in accuracy ($B = 0.91$, $SE = 2.18$, $z = 0.43$, $p = .67$) or looks, $F1(1, 31) = 0.41$, $p = .53$; $F2(1, 31) = 1.15$, $p = .29$. Comparing pointing accuracy between adults and children showed a large advantage for adults ($B = 2.68$, $SE = 0.48$, $z = 5.54$, $p < .0001$; the interaction slope for words was removed due to singular model convergence).

Discussion

This experiment asked whether 3- to 5-year-old children can learn minimal-pair words (*deev, teev*) as labels for different pictured objects. As a group they can, but they were outmatched by adult learners (60% accuracy vs. 94% accuracy). Not only was pointing accuracy low, but proportion looks to correct pictures also was low, suggesting that a more implicit measure like that used with infants (e.g., Yoshida et al., 2009) does not reveal stronger effects. This modest performance suggests that, as in the protracted perceptual learning hypothesis, children are still improving their skills at learning minimal-pair words even though they have roughly triple the language experience of 17-month-olds in studies by Werker and colleagues (e.g., Werker et al., 2002).

[2] As with accuracy data, if the 2 children who reversed responses are removed, consonant-pair looks become statistically significant, $t1(29) = 2.44$, $p = .02$; $t2(31) = 2.41$, $p = .02$.

[3] One learner in each condition completely reversed responses, without which accuracy would have been .97.

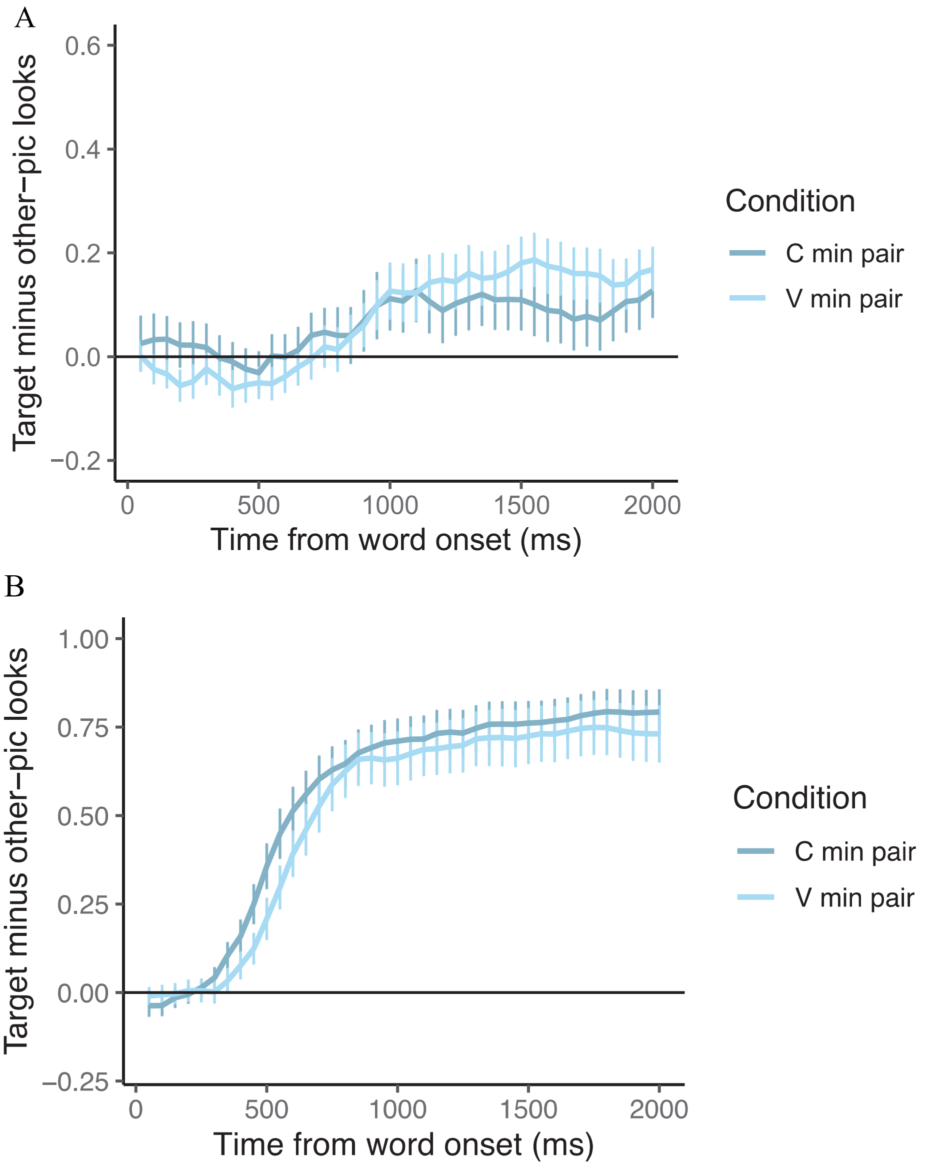

Fig. 3. Experiment 1: Children's (A) and adults' (B) target advantage (target looks minus looks to incorrect picture [pic]), with standard errors. Note that the *y* axes differ. C, consonant; V, vowel; min, minimal.

A secondary question in our experiment related to the roles of consonants versus vowels in word learning. We did not find pointing accuracy or visual fixation differences between different-consonant word pairs and different-vowel word pairs, and the small numerical difference was in favor of vowel differences being more salient. These findings are not consistent with previous research (Bonatti et al., 2005; Nazzi, 2005; Nazzi et al., 2016; Nespor et al., 2003), which has suggested that consonants are more indicative of word identity than vowels from an early point in development. It is not clear why our findings diverge from those of other researchers, but it may be related to differences in test language and perhaps test paradigm (see Nazzi et al., 2016, for a concise review of both). We return to this point in the General Discussion.

In short, our findings suggest that children aged 3 to 5 years still have more difficulty in learning novel minimal-pair words than adults. Nonetheless, we deemed it prudent to replicate in order to establish the stability of the effect and to rule out several counter-explanations. First, the finding of weak minimal-pair learning would be more convincing in the presence of strong learning of words that are not similar, that is, a high-performance control condition, to verify that participating children can learn words well without the challenge of phonological similarity. Second, the stimuli may have been less than optimal for child learning; they were single tokens of male adult-directed speech spoken in isolation. Children may have found isolated word labeling unnatural (see Fennell & Waxman, 2010, for evidence that 14-month-olds learn minimal pairs better from sentence contexts). Furthermore, children may be more accustomed to female caregiver voices and child-directed speech (but see Floccia et al., 2016, for limitations on the effectiveness of child-directed speech). Finally, there is evidence that 14-month-olds learn minimal-pair words better with high within-speaker token variability (Galle et al., 2014). Experiment 2 addressed these concerns.

Experiment 2

Here, we replicated Experiment 1, but instead of a male talker using a single token of each word in isolation in adult-directed speech, we recorded a female talker producing multiple word tokens in full sentences in child-directed speech. We also included a within-participants dissimilar word-learning condition to verify that children can perform to high levels in the word-learning task itself.

Method

Participants

We tested a new sample of 48 monolingual English-speaking children (21 girls), 16 each at 3, 4, and 5 years of age (M = 4.34 years, SD = 0.81). We chose the sample size of 48, slightly larger than that in the previous experiment, to provide a better age-balanced sample while also including each word pair equally often in each age group. (See Experiment 1 "Participants" section for power calculations.) An additional 2 children were tested but replaced due to refusal to point to named pictures (n = 1) or computer error (n = 1).

Stimuli

The words from Experiment 1 were recorded by a new speaker, a female young adult speaker from the geographical region of our university in the western United States, in child-directed speech in sentence frames (Table 3). Each sentence was recorded multiple times. Recordings were edited to select those free of disfluencies or other sound artifacts. Then stimuli were normalized to 70 dB SPL in Praat (Boersma & Weenink, 2014).

Procedure

As in Experiment 1, there were repeated learn–test cycles (Table 4), with some changes. As in Experiment 1, each block contained learning and test trials for a single pair of words. One change was that, in addition to a consonant minimal-pair block and a vowel minimal-pair block, there was a dissimilar word block (e.g., *dayge, fush*). This allowed within-participants assessment of ease of word learning when words were versus were not phonologically similar to each other. The second change was a reduction in the number of test trials from 16 to 8 per test. We did this to allow time for three learning–test cycles and because simulations with existing data suggested that an additional 8 trials did not provide substantially more information than the first 8 trials. A third change was that each participant received an additional fourth test phase following the third learning–test cycle that retested the dissimilar word pair. Because this fourth test phase answered a methodological question, we describe it in Appendix A.

Table 3
Experiment 2: Learning phase phrases and test phase phrases

Learning phase passages
Look at the X! Do you see the X?
The X! That's the X!
See the X over there? Isn't it a nice X?
It's a(n) X! Look at that X go!
Test phase sentences
Where's the X?
Point to the X!
Can you show me the X?
Find the X!

Note. "X" is a placeholder for the particular novel word heard. Learning phase phrases (usually two sentences) included two instances of each novel word, and test phase sentences included a single instance.

Table 4
Experiment 2 procedure: Example block orders

Block		Set 1	Set 2	Set 3	Set 4
1	Learn	dayge (pic3) fush (pic4)	*beesh (pic1) peesh (pic2*	zoof (pic8) zoohf (pic7)	soof (pic5) soohf (pic6)
	Test	dayge (pic3) fush (pic4)	*beesh (pic1) peesh (pic2*	zoof (pic8) zoohf (pic7)	soof (pic5) soohf (pic6)
2	Learn	*beesh (pic1) peesh (pic2)*	zoof (pic8) zoohf (pic7)	dayge (pic3) fush (pic4)	*bish (pic3)* *pish (pic4)*
	Test	*beesh (pic1) peesh (pic2*	zoof (pic8) zoohf (pic7)	dayge (pic3) fush (pic4)	*bish (pic3)* *pish (pic4)*
3	Learn	zoof (pic8) zoohf (pic7)	dayge (pic3) fush (pic4)	*beesh (pic1) peesh (pic2*	vush (pic1) tayge (pic2)
	Test	zoof (pic8) zoohf (pic7)	dayge (pic3) fush (pic4)	*beesh (pic1) peesh (pic2*	vush (pic1) tayge (pic2)
4	Retest	dayge (pic3) fush (pic4)	dayge (pic3) fush (pic4)	dayge (pic3) fush (pic4)	vush (pic1) tayge (pic2)

Note. There were 16 learning trials per block and 8 test trials per block. Four of 48 possible stimulus sets are shown. The first three are different orderings of the same stimulus pairs. Italics shows different-consonant trials, roman (plain) font shows different-vowel trials, and underlining shows dissimilar word trials.

Each of four learning passages included two instances of each word and was used twice per word for a total of 16 exposures to each word. Each testing sentence occurred once each for each word. Order of consonant, vowel, and dissimilar learning–test phases was counterbalanced across participants. Assignments of pictures to words were also counterbalanced across participants. Within each age group, all consonant and vowel minimal pairs were tested equally often.

Results

Pointing accuracy

As in Experiment 1, we conducted a logistic regression on pointing accuracy (Fig. 4). As before, age (centered and scaled) was included as a factor. Because the current experiment included three levels of word type—dissimilar words, consonant minimal pairs, and vowel minimal pairs—we used Helmert contrasts. Helmert contrasts are one way of representing a qualitative predictor variable with more than two levels in a regression model. In the current case of three qualitative levels (dissimilar, consonant, and vowel), this means two orthogonal contrasts; one compares the first level (dissimilar

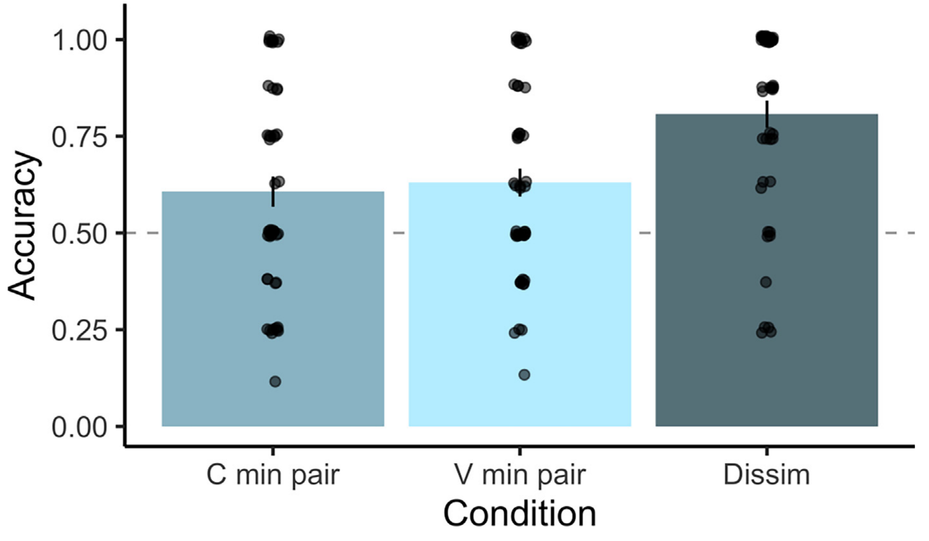

Fig. 4. Experiment 2: Children's pointing accuracy, with standard errors. Points are individual participants. C, consonant; V, vowel; min, minimal; Dissim, dissimilar.

words) with the mean of the next two levels (minimal pairs), and the other compares the next two levels with each other (consonant vs. vowel minimal pairs).

The intercept term was significant (B = 1.14, SE = 0.17, z = 6.82, p < .0001), indicating overall above-chance performance. The first word type contrast was significant (B = 1.56, SE = 0.37, z = 4.23, p < .0001), indicating that the dissimilar words (accuracy M = .807) were learned more accurately than the minimal-pair words (M = .617). However, the second word type contrast, which compared consonant and vowel minimal-pair learning, was not significant (B = 0.07, SE = 0.29, z = 0.22, p = .82). This indicates—as in Experiment 1—no detectable differences in accuracy between consonant and vowel minimal pairs. Considered individually, each contrast type exceeded chance: dissimilar words (B = 2.19, SE = 0.37, z = 5.95, p < .0001), consonant minimal pairs (B = 0.58, SE = 0.23, z = 2.57, p = .01), and vowel minimal pairs (B = 0.65, SE = 0.19, z = 3.52, p = .0004). However, both consonant and vowel minimal pairs showed relatively modest learning (.604 and .630, respectively).

The effect of age itself was not significant (B = 0.24, SE = 0.15, z = 1.56, p = .12), but there was a significant interaction of age with the second word type contrast (B = 0.54, SE = 0.27, z = 1.99, p = .047). To assess the interaction, we computed age effects for each word type individually. Consonant minimal pairs showed a significant increase with age (B = 0.55, SE = 0.20, z = 2.81, p = .005), whereas vowel minimal pairs showed no change with age (B = 0.02, SE = 0.18, z = 0.13, p = .90), nor did dissimilar words (B = 0.12, SE = 0.31, z = 0.38, p = .71).

Visual fixations

Data were preprocessed as before. ANOVAs were run on target advantage (target looks minus distractor looks; Fig. 5), collapsed within participants and words, respectively, with word type as a three-level predictor. The by-participants model also included the factor age. Word type was significant, $F1(2, 92)$ = 7.68, p = .0008; $F2(2, 62)$ = 9.45, p = .0003. Planned comparisons indicated that proportion looks to dissimilar words (.327 ± .259) exceeded looks to both consonant minimal-pair words (.136 ± .302), $t1(47)$ = 4.08, p = .0002; $t2(31)$ = 3.99, p = .0004, and vowel minimal-pair words (.145 ± .309), $t1(47)$ = 3.27, p = .002; $t2(31)$ = 3.23, p = .003. However, consonant and vowel minimal pair looks did not differ, $t1(47)$ = 0.15, p = .88; $t2(31)$ = 0.36, p = .72. Nonetheless, all three word types showed above-chance looks [dissimilar words: $t1(47)$ = 8.75, p < .0001; $t2(31)$ = 9.72, p < .0001; consonant minimal pairs: $t1(47)$ = 3.11, p = .003; $t2(31)$ = 4.21, p = .0002; vowel minimal pairs: $t1(47)$ = 3.25,

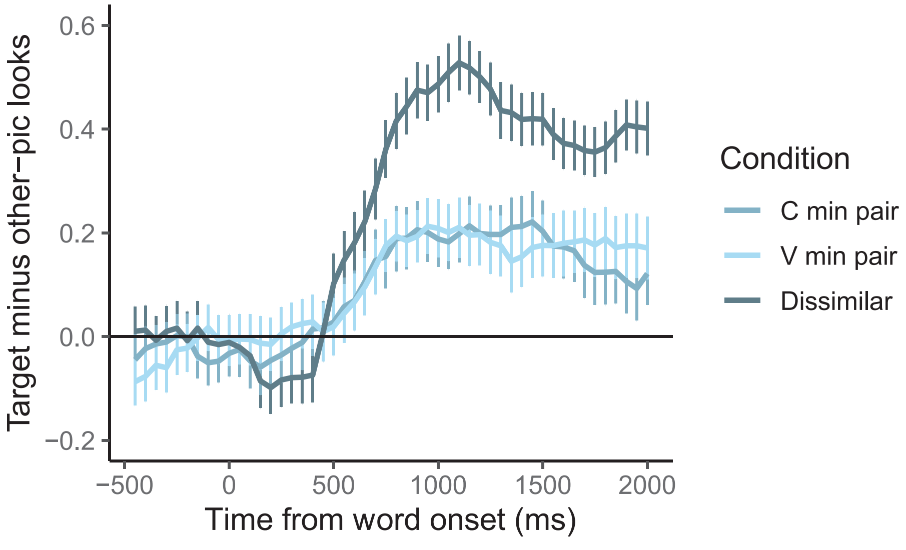

Fig. 5. Experiment 2: Children's fixations to target minus fixations to other picture (pic), with standard errors. C, consonant; V, vowel; min, minimal.

p = .002; *t2*(31) = 3.71, *p* = .0008]. The effect of age was significant, *F1*(1, 46) = 5.26, *p* = .03, indicating slightly greater target advantage from older children, but there was no interaction with word type, *F1* (2, 92) = 1.36, *p* = .26).

Discussion

The current experiment replicated Experiment 1 with a new child-directed talker, more natural sentence contexts rather than isolated words, and multiple word tokens rather than a single token per word. Despite all of these changes, results are strikingly similar to those in Experiment 1; minimal pairs were difficult, with pointing accuracy around 60%, and consonants did not show a learning advantage over vowels. This suggests that the findings of weak minimal-pair learning and equivalence of consonant and vowel contrasts are robust.

We also expanded on Experiment 1 by including a within-participant task control of dissimilar novel-word learning. When learning dissimilar words, children were significantly and substantially more accurate at close to 80%. This finding suggests that our word-learning task itself was not overly challenging but rather that minimal-pair distinctions increased the difficulty.

Age effects were mild overall and differed slightly between the two different dependent measures. Pointing accuracy improved with age for consonant minimal pairs but not for the other word pairs, whereas visual fixations suggested general improvements with age. Why the pointing accuracy age effects differed from those in Experiment 1, which showed marginal (pointing accuracy) and significant (looks) main effects of age, is not clear. This may be due to noise in the task itself, noise in the population of children tested, or both. In general, age effects are positive, which is consistent with gradual improvements in word-learning ability predicted by the protracted perceptual learning hypothesis.

We suggest that these results indicate that difficulty in learning similar-sounding words, which is present during late infancy (Tsui et al., 2019), persists through early childhood and likely well beyond. This is consistent with the protracted perceptual learning hypothesis. However, differences still exist between our test methodology and that of infant studies, mainly infant switch studies. In those studies, each word is repeated far more times (e.g., 64 times per word; Rost & McMurray, 2009, 2010; Stager & Werker, 1997; Yoshida et al., 2009) than was the case here (16 exposures per word). Whether this massive number of exposures is actually a causal factor in infants' learning is debatable given evidence in even younger children (2-year-olds) that massed exposures to word-learning instances are not as successful as distributed exposure (Vlach et al., 2012). Still, given that this is one striking difference between infant paradigms and the current paradigm, and because it is intuitively reasonable that additional learning might increase knowledge, Experiment 3 assessed whether additional word exposure facilitates minimal-pair learning.

Experiment 3

This study tested whether the youngest set of children in our age range can learn minimal pairs more effectively with a higher number of word exposures, like that seen in dishabituation studies where 17-month-old infants show evidence of learning. The study was modeled after that of Yoshida et al. (2009), which is the one example in the literature of eye-tracked recognition testing of infants' similar-sounding word learning. Those authors used a habituation phase like earlier tasks (Stager & Werker, 1997), but like the current Experiments 1 and 2 they measured learning by testing visual fixations to the named picture.

Note that our goal was *not* to conduct a habituation study with 3-year-olds but rather to increase the amount of word exposure to something like that seen in habituation studies. Thus, we decided to take a reasonable average number of word exposures and trials across numerous infant habituation studies (~64 instances per word, a quadrupling of that in Experiments 1 and 2) and to present that number to each child across an exact number of trials (16). If heightened word exposure is the critical variable allowing infants to succeed at minimal-pair learning, then 3-year-olds should succeed here. However, if 3-year-olds need many more years of language experience to achieve fluency in minimal-pair learning, as in the protracted perceptual learning hypothesis, then 3-year-olds should find learning in this task to be difficult, as they did in Experiments 1 and 2.

We also included a post-test with two highly familiar minimal pairs to verify that children can indeed distinguish familiar minimal pairs as well as to explore relationships between learning ability and familiar minimal-pair recognition. On the view that the objects of learning are individual speech sounds, children who better distinguish familiar minimal pairs should also better learn novel minimal-pair words because both of these reflect better speech sound knowledge. The protracted perceptual learning hypothesis, however, predicts that knowledge of familiar words reflects word-specific knowledge, with limited predictive power for ability to learn novel words.

Method

Participants

We tested 24 3-year-olds (M = 3.56 years, SE = 0.28; 6 girls), all from monolingual English-speaking homes. (See Experiment 1 "Participants" section for power calculations.) An additional 4 participants were tested but not included due to exposure to another language (n = 1), opting to leave the experiment (n = 2), or taking off headphones (n = 1).

Stimuli

We used a subset of the word stimuli from Experiment 2, specifically the different-consonant word pairs that contained stop consonants and tense vowels: *deev, teev; dayge, tayge; beesh, peesh;* and *boove, poove.* Stop consonants have been used in a variety of switch word-learning studies with infants (e.g., Galle et al., 2014; Rost & McMurray, 2009, 2010; Stager & Werker, 1997; Yoshida et al., 2009). Furthermore, Treiman et al. (1998) found that 5-year-olds differentiated stop voicing contrasts slightly better than fricative voicing contrasts, so using only the stop voicing contrasts should in principle work to children's advantage in the current study. Words for exposure were excised from a superset of Experiment 2 learning phase sentence recordings and combined to produce a prosodically varied list of repetitions of each word. Words from sentence contexts where coarticulation yielded a noncanonical pronunciation were not used. Eight different random orders of the same set of 8 tokens were created in Praat 5.4.01 (Boersma & Weenink, 2014). Each random order was written to a single sound file and presented on a different learning trial. Within each of these recordings, word onsets were spaced 2000 ms apart, so that the entire recording lasted 16 s, including a final trailing silence. Word tokens used at test were 4 additional tokens excised from Experiment 2's test phase declarative sentences. Recordings of real words for the post-test were obtained from a variety of sources, mostly previous studies in our lab. Like the novel-word talker, all familiar-word talkers were female.

Procedure

We followed Yoshida et al. (2009) in several respects because they found evidence of minimal-pair learning in an eye-tracked recognition test in 14-month-olds. We borrowed other structural elements from Rost and McMurray (2009, 2010), as described below. Briefly, Yoshida et al. (2009) presented two novel pictures, each with a novel label (*bin* or *din*), in a habituation phase. On each habituation trial, children heard up to 10 instances of each word. The average number of trials to habituation in Yoshida et al. (2009) was 12.67, meaning that the *maximum* possible average number of tokens per word was about 56 words,[4] similar to our exposure rate of 64 instances per word (8 learning trials for each word × 8 words per trial; see Rost & McMurray, 2009, 2010, for similar numbers). Similar to Rost and McMurray (2009, 2010; see also Galle et al., 2014), we allowed trials to continue to completion even if children looked away. This could contribute to sound familiarity even if children were not fixating the display. It is also similar to Experiments 1 and 2, where children were directed but not required to look at the screen during learning.

Following 16 learning trials, we presented a test phase. Like Yoshida et al. (2009), we presented interspersed real-word test trials using the pairs *dog, baby* and *car, shoe* to encourage children to treat the task as a word recognition task, and we made sure to present at least two real-word test trials before the novel-word test trials began. We included an 8-trial post-test of minimal-pair words that should be highly familiar to young children: *doll* and *ball* and *boot* and *boat*.[5] This post-test assessed whether children were capable of distinguishing at least some highly familiar minimal pairs. It also served to verify that children were still paying attention at the end of the study.

Results

Novel words

Pointing accuracy. Throughout these results, pointing accuracy (Fig. 6) was assessed with mixed-effects logistic regression models with an intercept term (which assessed above-chance performance) and participants and words random intercepts. Novel-word pointing accuracy was .55 (SD = .18), which did not exceed chance (B = .21, SE = .15, z = 1.43, p = .15). It was also comparable to stop-consonant minimal-pair accuracy of the 3-year-olds in Experiment 1 (.55, n = 4) and Experiment 2 (.59, n = 8) (see Fig. 7 for all pointing accuracy results from 3-year-olds). There was no relationship between novel-word pointing accuracy and age, $r(22)$ = .19, p = .38) over this restricted age range.

Visual fixations (to novel pictures). Did performance on total visual fixations to novel pictures reveal similarities to Yoshida et al. (2009) that pointing accuracy obscured? No. Target advantage (Fig. 8) was −.002 (SD = .197; 10 of 24 children exceeded 0). As with pointing accuracy, visual fixations did not exceed chance, $t1(23)$ = 0.04, p = .97; $t2(7)$ = 0.13, p = .90. The correlation of target advantage with age did not approach significance, $r(22)$ = .12, p = .56.

[4] This assumes that all infants listened for the maximum trial duration (10 instances) on all but the last 4 trials, where they listened to exactly 65% (the maximum listening duration that would still have met the habituation criterion). With a mean of 12.67 trials to habituation, 8.67 trials * 10 words * 100% looking time + 4 trials * 10 words * 65% looking time = 112.7 words. Dividing this by 2 yields 56.35 heard instances of each of the two words. It is unlikely that all infants listened for the entire trial prior to habituation, so this is probably an overestimate.

[5] Some minimal pairs that better match the consonant contrasts in our novel words (e.g., *beach, peach; goat, coat*) in Experiments 1 and 2 were rejected because we were not confident that 3-year-olds would readily recognize both words.

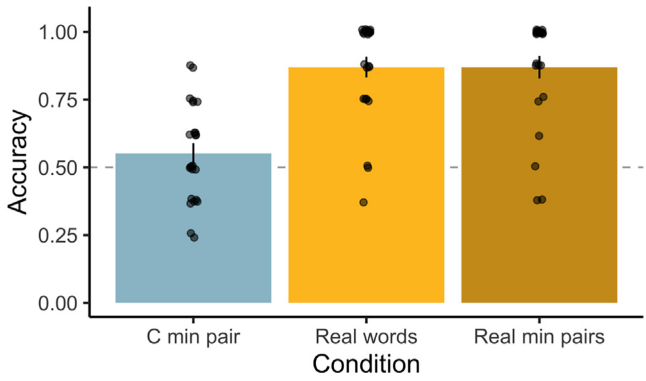

Fig. 6. Experiment 3: Children's pointing accuracy, with standard errors. C, consonant; min, minimal.

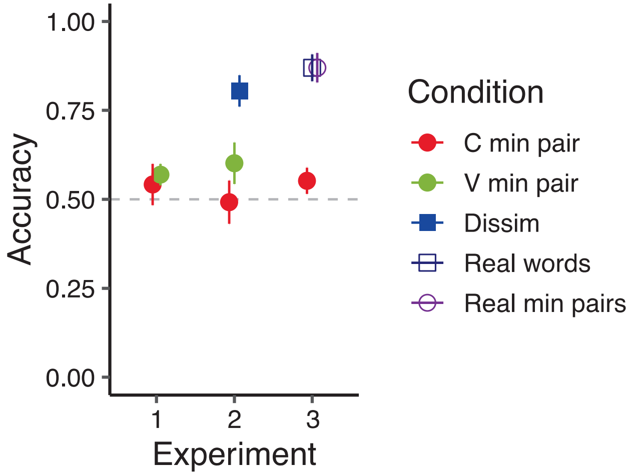

Fig. 7. Three-year-olds' pointing accuracy across experiments. C, consonant; V, vowel; min, minimal; Dissim, dissimilar.

Familiar words

Pointing accuracy. Unlike novel words, familiar-word pointing accuracy during the test was high (M = .87, SD = .19), exceeding chance (B = 2.64, SE = 0.54, z = 4.89, p < .0001). This suggests that children were overall attentive during the test. The correlation with age was not significant, $r(22) = -.09$, p = .67. The correlation between accuracy on these familiar words and accuracy on novel words was marginally negative, $r(22) = -.38$, p = .07, opposite the expected direction if better familiar-word knowledge predicts novel-word learning.

Visual fixations. As with the pointing accuracy results for familiar words, looks exceeded chance, $t1(23)$ = 6.45, p < .0001; $t2(3)$ = 14.62, p = .0007, with a target advantage of .24 (SD = .18; 20 of 24 children exceeded 0). The correlation with age did not approach significance, $r(22)$ = .12, p = .58, nor did the correlation between familiar-word target advantage and novel-word target advantage, $r(22)$ = .21, p = .33.

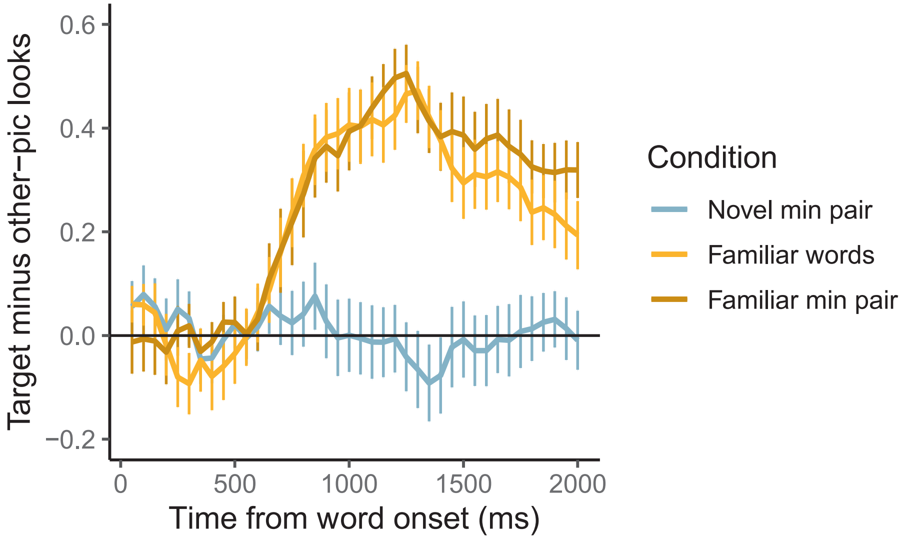

Fig. 8. Experiment 3: Children's fixations to target minus fixations to other picture (pic), with standard errors. min, minimal.

Familiar minimal-pair post-test

Pointing accuracy. Pointing accuracy on the minimal-pair post-test was high at .87 (SD = .20), which exceeded chance (B = 4.06, SE = 1.36, z = 2.99, p = .003). Furthermore, it verified both that children were still attentive by the end of the test and that the children we tested were capable of differentiating highly familiar minimal-pair words. The correlation with age was not significant, $r(22) = -.13$, p = .53, nor was the correlation between accuracy on these familiar minimal pairs and learning accuracy, $r(22) = -.05$, p = .82.

Visual fixations. Fixations patterned with pointing accuracy; looks exceeded chance, $t1(23) = 9.12$, $p <$ .0001; $t2(3) = 3.24$, p = .048, with a target advantage of .28 (SD = .15; all 24 children exceeded 0). The correlation of target advantage with age did not approach significance, $r(22) = .09$, p = .66. The correlation between target advantage for these familiar minimal pairs and target advantage for novel minimal pairs was not significant, $r(22) = .26$, p = .21.

Discussion

We asked whether a larger "dosage" of words, more comparable to that used in infant studies of minimal-pair word learning, might lead to better learning in our age range. It did not; despite quadrupling the number of word instances heard, 3-year-old children did not reach above-chance pointing accuracy or looking proportions and did not show higher accuracy than same-age children who heard fewer word tokens (Experiments 1 and 2). Nonetheless, the same children showed excellent pointing identification and visual fixations to familiar words, including two familiar-word minimal pairs. This indicates that they were attentive and had the capacity to make fine-grained phonetic distinctions. Yet this phonetic capacity did not relate to remembering novel minimal pairs.

General discussion

Consistent with the protracted perceptual learning hypothesis, we found that preschool-aged children show significant difficulty in learning new minimal-pair words. This held in a set of 32 different word pairs (16 consonant-differing word pairs and 16 vowel-differing word pairs), two different talkers, two different levels of exposure, and two dependent measures. In general, performance improved with age. This pattern of results suggests that minimal-pair learning success in 17-month-olds (e.g., Stager & Werker, 1997; Werker et al., 2002) reflects the beginnings of a learning process that continues for years.

We also asked whether there were differences between consonant minimal-pair learning (*deev, teev*) and vowel minimal-pair learning (*deev, div*). According to accounts positing greater consonant sensitivity, consonant-differing words should be more distinct, and therefore easier to learn, than vowel-differing words (e.g., Nazzi et al., 2016; Nespor et al., 2003). This is not what we found. In both Experiments 1 and 2, children performed with similar accuracy on consonant-differing and vowel-differing pairs.

Theoretical implications

Our findings do not support early stabilization of speech sound categories during late infancy. If categories stabilized within the first few years of life, children aged 3 to 6 years should not experience substantial difficulty in learning word pairs distinguished by these speech sound categories or should show marked improvement within this time range. Our findings are more consistent with a protracted perceptual learning process, with improvements presumably taking place between 5 or 6 years (the oldest age tested) and young adulthood. On this account, tasks commonly used to test infant sound pattern knowledge, such as habituation/dishabituation (which includes the switch procedure), reflect the beginnings of learning a language's sound patterns. Thus, our results suggest that the wealth of infant speech perception and word-learning studies should be interpreted as the beginnings of a lengthy speech sound learning process. Of course, a remaining construal of these results is that similar things are harder to learn or harder to differentiate throughout development, with weaker learners (children) showing these difficulties more strongly. That is, children have *overall* weaker memory skills (not just within linguistic material), and this leads them to generally slower learning of similar-sounding words. The current data cannot dismiss this explanation, although future studies might (see "Limitations ..." section below).

A related question here is, assuming that there is protracted development of word learning, at what age do children reach adult-like performance? We suspect that it may be 12 years or later, based on findings like Hazan and Barrett (2000) and McMurray et al. (2018). In both of these studies, children as old as 12 years performed in a non-adult-like manner, suggesting more diffuse speech sound or word form representations. We expect that representations become slowly richer as learners experience a wide array of phonetic variability (e.g., hearing words spoken by different speakers, in different sentence contexts, and in different affects; hearing sounds across different words). These richer representations in turn better support learning of novel but related forms. This account, by specifying perceptual learning as a mechanism, also makes predictions about adults' abilities to learn similar-sounding words. We return to this point later.

A U-function in speech pattern learning between infancy and adulthood? Probably not

A curious pattern is that whereas our 3-year-olds appeared not to learn minimal-pair words in Experiment 3, 17-month-old infants in many other studies did (see Introduction). At a superficial level, one might infer a U-shaped pattern in which infants have good sound pattern knowledge, which then weakens during preschool and is slowly regained toward young adulthood. However, this account violates parsimony: it is at odds with a continuously increasing, or even nondecreasing, picture of developmental change. It also glosses over substantial differences between research paradigms used with infants versus young children and adults. These differences between paradigms are a major obstacle in connecting findings across age ranges (see Creel & Quam, 2015). The current study, especially Experiment 3, is a step in that direction, but more work needs to be done.

Across various switch studies (see Tsui et al., 2019), infants at 14 to 17 months of age sometimes showed significant minimal-pair recognition, but our 3-year-olds in Experiment 3 did not. Why? One possible explanation along the lines of PRIMIR (Werker & Curtin, 2005) is that the developmental level of the task is more appropriate for the attentional filters of a 14-month-old than for a 3-year-old; perhaps 3-year-olds would be more engaged with a narrative task. Yet another possible explanation is the testing environment. Infants typically are tested in darkened rooms with people and objects out of view, and a large display is the brightest, most interesting item in the room. Our 3-year-olds were tested in preschools, which may contain interesting objects (books, toys, and experimenters) with a computer monitor of typical size. Admittedly, neither of these explanations is fully satisfying. In the "Limitations ..." section below, we outline some future approaches that might delineate change over time more clearly.

Lack of consonant bias in word learning

The current work suggests that children find consonant minimal pairs and vowel minimal pairs roughly equally difficult to learn. Yet much previous work (Bonatti et al., 2005; Nazzi, 2005; Nazzi et al., 2016; Nespor et al., 2003) suggests a stronger weighting of consonants than vowels in word recognition, beginning sometime around 12 months of age and continuing into adulthood (Nazzi et al., 2016; New et al., 2008; see also Creel et al., 2006, for related work in adults). Why the discrepancy between the current study and earlier findings? This might be a language difference in that much of Nazzi and colleagues' work was done in France with French speakers. Previous studies of (British) English-learning children are more mixed than French findings, with two of three studies suggesting no consonant/vowel differences during infancy and toddlerhood (Floccia et al., 2014; Mani & Plunkett, 2007; but Nazzi et al., 2009, found differences). Similarly, Creel (2012) and Swingley (2016) reported equivalent vowel and consonant sensitivity in mispronunciations of familiar words in American English. Thus, the current study is consistent with literature on English-speaking children's (lack of) consonant biases and may suggest that consonant biases seen in English adults emerge more slowly during development than consonant biases in other languages.

In any case, it is important to bear in mind that comparing consonant similarity with vowel similarity is a complex problem. Although there are modes of quantifying word and speech sound similarity (such as articulatory feature distance, acoustic–phonetic feature distance, similarity ratings, confusability in noise, memory errors), there is currently not an ideal way to determine whether a particular consonant pair is equivalent in perceptual distance to a particular vowel pair. One might take the current findings to indicate that at least some vowel pairs are equivalent to at least some consonant pairs in distinguishing words in English.

Limitations and a proposal for exploring continuity across development

We have suggested that our findings reflect persistent difficulty in learning minimal-pair words. One limitation is that our word-learning studies do not perfectly mimic most word-learning studies on infants and toddlers (e.g., Stager & Werker, 1997). As noted earlier, the studies presented here differ in terms of task characteristics, exact word pairs used, and distractions available to infants versus our preschool participants. Experiments 1 and 2 also presented far fewer word exposures than those seen in many infant studies. We equated this factor in Experiment 3, but learning did not improve.

An additional limitation, mentioned above, is that results are also consistent with a simpler account that similar things are harder to learn than less similar things, combined with general developmental improvements in learning, rather than a *selective* improvement over development in similar-sounding word learning. Future approaches might more fruitfully distinguish between these two accounts, as we outline below in our recommendations for future work connecting across development.

One recommendation is that researchers who examine age-related changes in similar-word learning ability in both infant and young-child age groups should use a fixed amount of exposure and an eye-tracked test phase. For examining a wider age range, such as infancy through adulthood or early childhood through adulthood, we recommend testing under matched task conditions, matched difficulty, and the same set of stimuli. That raises the question of how to match difficulty; do learners of all ages get equivalent amounts of exposure, or should exposure be titrated by age group? Here, we chose to match exposure, which put adult learners at ceiling when learning two words at a time. Some outside the field of language research might regard this as an interesting finding on its own given that it contradicts the simplified story that children are better language learners than adults in all respects (see, e.g., Genesee, 1981, Krashen et al., 1979, and Snow & Hoefnagel-Höhle, 1978, for discussion of better vocabulary learning in older children and adults than in young children). Still, it seems likely that this is an unfair comparison given children's reduced attentional and working memory capacities relative to those of adults. Thus, a different possibility[6] would be to determine the amount of exposure needed to reach a particular accuracy criterion at each age on dissimilar word learning and then deliver that amount of exposure to similar-sounding words. A caveat is that quadrupling exposure did not appear to improve 3-year-olds' word learning in Experiment 3 versus Experiments 1 and 2, but perhaps *decreasing* exposure for adults, or inserting a long delay between learning and test, could pull older learners off ceiling.

Another possibility would be to give older learners a larger number of words to learn than children, which clearly lowers adult performance in learning similar words (e.g., Pająk et al., 2016; White et al., 2013). Of course, giving adults more words to learn might make the task harder in ways that change the task entirely, in particular by creating a larger set of recently learned competitors than is present for children learning a single pair of words. More competitors increases confusion and lowers performance, but then it would be hard to know whether we had appropriately calibrated task difficulty or simply provided a larger competitor (confusion) space. A middle ground might be to make the task slightly more complex than the child task used here, asking children and adults to learn, say, four or eight words at a time instead of two at a time and giving adults less exposure and more delayed testing.

A very different approach might address whether the proposed mechanism—lengthy perceptual learning—leads to similar outcomes despite chronological age. Thus, one might match child difficulty to adult difficulty by testing comparison adult learners with speech perceptual contrasts (or even nonspeech perceptual contrasts) that they are inexperienced with. If the critical difference in performance between children and adults is cognitive (im)maturity, then adults should excel despite perceptual inexperience. If the critical difference is instead perceptual familiarity, then adults should perform like children. This basic idea was partially tested by Pajak et al. (2016), who found that adults who were unfamiliar with subtle sound contrasts in another language showed low accuracy in a word-learning task (as well as a perceptual discrimination task). Future work could more carefully match perceptual discriminability in child and adult learners to set up comparable word-learning situations.

Our finding of difficult minimal-pair learning generalizes across 32 word pairs exemplifying 8 different minimal contrasts (see Appendix B for details). However, a limitation is that the finding of learning difficulty might not generalize perfectly to other *types* of contrasts. For example, it is possible that children would learn words better with consonant place or manner contrasts rather than consonant voicing contrasts (see Jusczyk et al., 1999) or other sorts of vowel contrasts besides tense–lax distinctions. It is also possible that consonant biases—better learning of consonant-differing words than of vowel-differing words—might be evident if using a different consonant feature. For that matter, consonant biases might appear to be quite different if one compared vowels with *coda* (end-of-syllable) consonants rather than syllable-onset consonants given that some literature suggests that coda consonants are weakly processed (Creel et al., 2006; Redford & Diehl, 1999; but see Nazzi & Bertoncini, 2009).

[6] Thanks go to an anonymous reviewer for suggesting these variations.

Related to the above, it is interesting that performance was not at ceiling even for dissimilar words. Although this might reflect children's imperfect attentional capacity, it might also be the case that our "dissimilar" words, all of which were closed (consonant-final) monosyllables, were similar enough to each other to cause mild confusion. Performance might improve further with words that differ in syllable type (open or closed), syllable number, or stress pattern.

A final limitation is a methodological one: Looking responses are partly obscured when pointing responses are also made. Recall that this is due to pointing movements sometimes obscuring the eye-tracking camera. Although we partly addressed this with a data extrapolation procedure, future work might obtain cleaner gaze data by using an eye-tracking modality that is not obscured by pointing movements, such as a head-mounted tracker, or by obtaining only looking responses on some trials and only pointing on others.

Conclusion

We found that preschool-aged children (3–6 years) have ongoing difficulty in encoding form–meaning mappings of similar-sounding words like *deev* and *teev,* even though the youngest of these children readily recognize familiar similar-sounding words like *doll* and *ball.* Findings are consistent with a protracted perceptual learning hypothesis that sound pattern encoding slowly increases in fidelity with increasing perceptual experience with particular words over a protracted time period, with phonetic regularities eventually emerging to support novel-word learning. However, other explanations remain possible. Accordingly, further work using more uniform dependent measures and matched stimuli is needed to map out continuous trajectories of learning ability across age.

CRediT authorship contribution statement

Sarah C. Creel: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Conor I. Frye:** Conceptualization, Data curation, Formal analysis, Software, Writing – original draft.

Data availability

Data are available at the Open Science Framework (https://doi.org/10.17605/OSF.IO/TKDXC).

Acknowledgments

Thanks go to Elizabeth Groves, Alicia Escobedo, and Creel lab research assistants for data collection. S.C.C. and C.I.F. were supported by National Science Foundation grants (BCS-1230003 and BCS-1057080).

Appendix A

Experiment 2 post-test performance

Experiment 2 presented a delayed retest of dissimilar word pairs to gauge whether knowledge was retained past the immediate post-learning time period. We opted to use the dissimilar word pair for retest on the logic that accuracy would be higher for these pairs, meaning that we would be better able to detect continued above-chance performance even if there was degradation by the time of the post-test.

We wanted to address a question about child word learning paradigms in general, namely the extent to which they reflect only fleeting memories of word–picture association. One might apply this critique both to our experiments and to the switch paradigm used with infants (Stager & Werker, 1997) or any number of child word-learning experiments that test immediately after learning. Specifically, in our case children might remember the mapping (e.g., Object 1 = "deev") from the last learning trial and perform accurately on the first test trial (if "deev", pick Object 1; otherwise, pick Object 2). On the next test trial, they might remember the mapping from the first test trial, and if the word is the same they pick the same picture; if it is different, they pick the other picture, and so on. This would reflect very short-term knowledge of the taught words (see Horst & Samuelson, 2008, on short-duration learning). In fact, we suspect that this is the reason why children in studies of this type occasionally reverse the responses (e.g., showing 5% accuracy). However, children's performance may also reflect formation of longer-term representations. If so, then knowledge should persist after a delay even if that delay is filled with learning other sets of words. To begin answering this question, we included a final post-test in which we retested the dissimilar word pair to assess retention of knowledge once this short-term strategy was no longer available.

Before analyzing post-test accuracy, we dropped participants whose post-test immediately followed their dissimilar word train–test cycle because these participants could still have been using association memory from the end of the previous test. Of the remaining 32 participants, 6 were missing the post-test due to a programming error. This left 26 participants. Accuracy was .67 (SD = .32), which was significantly lower than the same children's initial dissimilar word performance (M = .83, SD = .23; B = 1.41, SE = 0.64, z = 2.21, p = .03) but nonetheless exceeded chance (B = 1.15, SE = 0.43, z = 2.71, p = .007). Looks (Fig. A1) were greater to dissimilar words during the main test (.405 ± .266) than during the post-test (.165 ± .345), $F1(1, 25)$ = 13.25, p = .001; $F2(1, 31)$ = 13.80, p = .0008. However, looks during the post-test still exceeded chance, $t1(25)$ = 2.44, p = .02; $t2(31)$ = 2.58, p = .01. These results suggest that children were not simply using short-term memory strategies to perform well in the test phase but in fact were encoding longer-term word forms and associations. Still, the drop in performance over a relatively short time period suggests that immediate testing does provide an especially rosy picture of word learning.

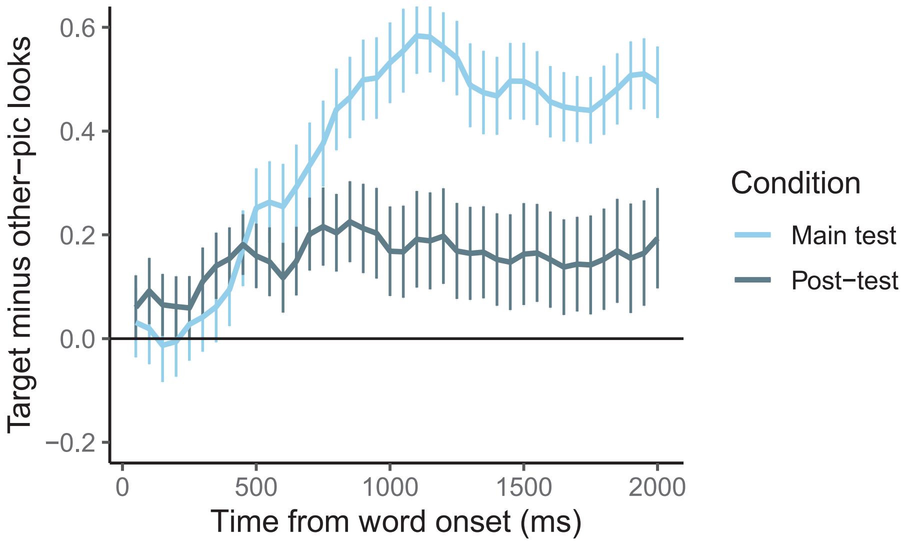

Fig. A1. Looks to dissimilar named pictures in Experiment 2 main test (immediately after learning) and post-test (after at least one other learning–test cycle intervened). pic, picture.

Appendix B

Pointing accuracy of each of the word pairs from Experiments 1 and 2. Fig. B1

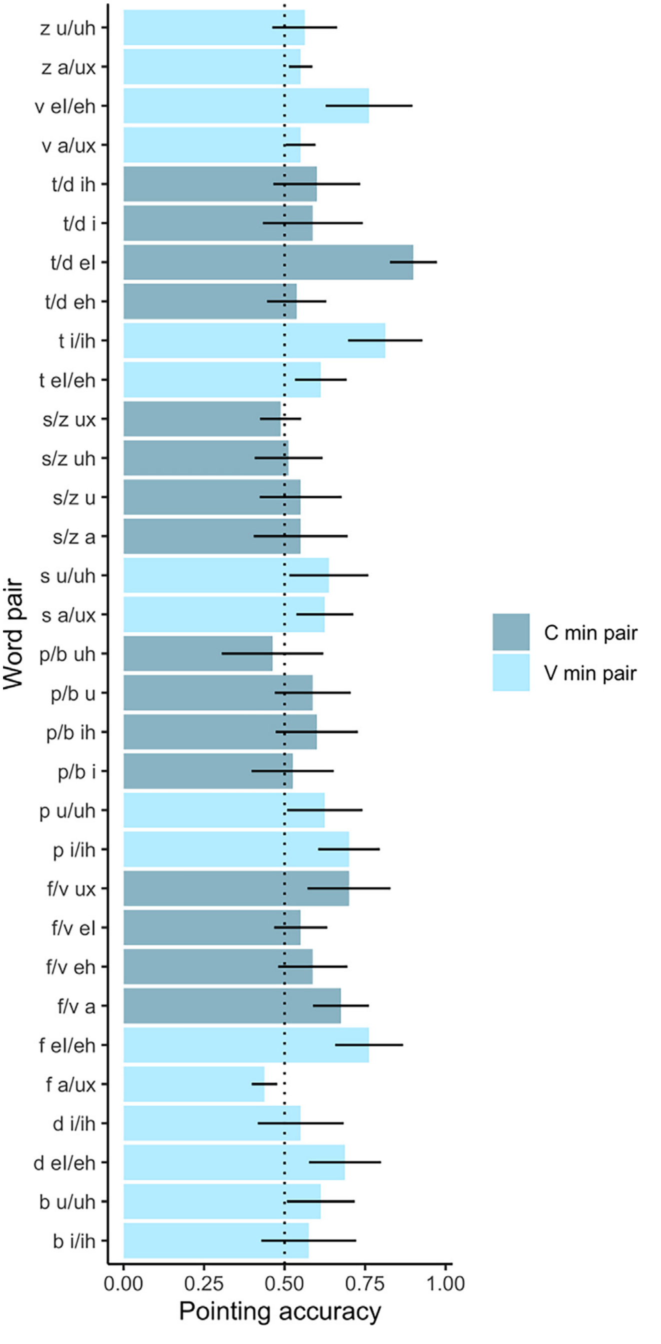

Fig. B1. Accuracy of the different word pairs combined across Experiments 1 and 2. There are 5 participants per bar. Darker bars are consonant word pairs (such as *deev*/*teev*, labeled as t/d i), and lighter bars are vowel word pairs (such as *deev*/*div*, labeled as d i/ih). C, consonant; V, vowel; min, minimal.

References

Adolph, K. E., Vereijken, B., & Shrout, P. E. (2003). What changes in infant walking and why. *Child Development, 74*(2), 475–497. https://doi.org/10.1111/1467-8624.7402011.

Anzures, G., Mondloch, C. J., & Lackner, C. (2009). Face adaptation and attractiveness aftereffects in 8-year-olds and adults. *Child Development, 80*(1), 178–191. https://doi.org/10.1111/j.1467-8624.2008.01253.x.

Barton, D. (1976). Phonemic discrimination and the knowledge of words in children under 3 years. *Papers and Reports on Child Language Development, 11,* 61–68. (Distributed by ERIC Clearinghouse)

Bates, D., Maechler, 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.

Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer* [computer program]. Version 5.4.01. Retrieved November 9, 2014, from http://www.praat.org

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations. *Psychological Science, 16* (6), 451–459 http://www.jstor.org/stable/40064248.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357.

Bril, B., & Ledebt, A. (1998). Head coordination as a means to assist sensory integration in learning to walk. *Neuroscience and Biobehavioral Reviews, 22*(4), 555–563. https://doi.org/10.1016/S0149-7634(97)00044-4.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The EyeLink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox Retrieved from. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617 http://www.ncbi.nlm.nih.gov/pubmed/12564564.

Creel, S. C. (2012). Phonological similarity and mutual exclusivity: On-line recognition of atypical pronunciations in 3–5-year-olds. *Developmental Science, 15*(5), 697–713. https://doi.org/10.1111/j.1467-7687.2012.01173.x.

Creel, S. C. (2014). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language, 73*, 81–98 https://doi.org/10.1016/j.jml.2014.03.001.

Creel, S. C. (2018). Accent detection and social cognition: Evidence of protracted learning. *Developmental Science, 21*(2), Article e12524. http://doi.org/10.1111/desc.12524

Creel, S. C. (2022). Preschoolers have difficulty discriminating novel minimal-pair words. *Journal of Speech Language and Hearing Research, 65*, 2540–2553. https://doi.org/10.1044/2022_JSLHR-22-00029.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language, 54,* 1–19. ttps://doi.org/10.1016/j.jml.2005.09.003

Creel, S. C., & Dahan, D. (2010). The effect of the temporal structure of spoken words on paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 110–122. https://doi.org/10.1037/a0017527.

Creel, S. C., & Jiménez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology, 113*, 487–509. https://doi.org/10.1016/j.jecp.2012.07.007.

Creel, S. C., & Quam, C. (2015). Apples and oranges: Developmental discontinuities in processing? *Trends in Cognitive Sciences, 19* (12), 713–716. https://doi.org/10.1016/j.tics.2015.09.006.

Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology, 105* (4), 376–385. https://doi.org/10.1016/j.jecp.2009.12.004.

Curtin, S., Fennell, C. T., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word–object associative learning. *Developmental Science, 12*(5), 725–731. https://doi.org/10.1111/j.1467-7687.2009.00814.x.

Escudero, P., Mulak, K. E., & Vlach, H. A. (2016). Cross-situational learning of minimal word pairs. *Cognitive Science, 40*(2), 455–465. https://doi.org/10.1111/cogs.12243.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Fecher, N., & Johnson, E. K. (2018). Effects of language experience and task demands on talker recognition by children and adults. *Journal of the Acoustical Society of America, 143*(4), 2409–2418. https://doi.org/10.1121/1.5032199.

Fennell, C. T., & Byers-Heinlein, K. (2014). You sound like mommy: Bilingual and monolingual infants learn words best from speakers typical of their language environments. *International Journal of Behavioral Development, 38*(4), 309–316. https://doi.org/10.1177/0165025414530631.

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.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech, 46*(Pt 2–3), 245–264. https://doi.org/10.1177/00238309030460020901.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research, 38*, 379–412. https://doi.org/10.1007/s10936-008-9097-8.

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., ... Vihman, M. M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition, 148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004.

Floccia, C., Nazzi, T., Delle Luche, C., Poltrock, S., & Goslin, J. (2014). English-learning one- to two-year-olds do not show a consonant bias in word learning. *Journal of Child Language, 41*(5), 1085–1114. https://doi.org/10.1017/S0305000913000287.

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2014). The role of single talker acoustic variation in early word learning. *Language Learning and Development, 11*(1), 66–79. https://doi.org/10.1080/15475441.2014.895249.

Garnica, O. K. (1971). The development of the perception of phonemic differences in initial consonants by English-speaking children: A pilot study. *Papers and Reports on Child Language Development, 3*, 1–31 https://files.eric.ed.gov/fulltext/ED111173.pdf.

Genesee, F. (1981). A comparison of early and late second language learning. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement, 13*(2), 115–128. https://doi.org/10.1037/h0081168.

Gerken, L., Murphy, W. D., & Aslin, R. N. (1995). Three- and four-year-olds' perceptual confusions for spoken words. *Perception & Psychophysics, 57*(4), 475–486. https://doi.org/10.3758/BF03213073.

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition, 118*(2), 201–210. https://doi.org/10.1016/j.cognition.2010.11.002.
Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *British Journal of Developmental Psychology, 26*(3), 409–433. https://doi.org/10.1348/026151007X251712.
Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language, 17*(1), 171–183. https://doi.org/10.1017/S0305000900013167.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1166–1183. https://doi.org/10.1037/0278-7393.22.5.1166.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251.
Havy, M., Bertoncini, J., & Nazzi, T. (2011). Word learning and phonetic processing in preschool-age children. *Journal of Experimental Child Psychology, 108*(1), 25–43. https://doi.org/10.1016/j.jecp.2010.08.002.
Havy, M. M., Serres, J., & Nazzi, T. (2014). A consonant/vowel asymmetry in word-form processing: Evidence in childhood and in adulthood. *Language and Speech, 57*(2), 254–281. https://doi.org/10.1177/0023830913507693.
Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28*(4), 377–396. https://doi.org/10.1006/jpho.2000.0121.
Højen, A., & Nazzi, T. (2016). Vowel bias in Danish word-learning: Processing biases are language-specific. *Developmental Science, 19*(1), 41–49. https://doi.org/10.1111/desc.12286.
Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*(2), 128–157. https://doi.org/10.1080/15250(xH)701795598.
Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–165). Academic Press.
Jusczyk, P. W., Goodman, M. B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language, 40*, 62–82. https://doi.org/10.1006/jmla.1998.2605.
Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science, 18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x.
Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly, 13*(4), 573–582. https://doi.org/10.2307/3586451.
Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202.
Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language, 57*(2), 252–272. https://doi.org/10.1016/j.jml.2007.03.005.
McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472–1491. https://doi.org/10.1037/dev0000542.
McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science, 317*(3 August), 631.
Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Hearing Research, 48*, 1033–1047. https://doi.org/10.1044/1092-4388(2005/072).
Munson, B., Swenson, C. L., & Manthei, S. C. (2005). Lexical and phonological organization in children: Evidence from repetition tasks. *Journal of Speech, Language, and Hearing Research, 48*(1), 108–124. https://doi.org/10.1044/1092-4388(2005/009).
Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition, 98*(1), 13–30. https://doi.org/10.1016/j.cognition.2004.10.005.
Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: New evidence from consonants in coda positions. *Language and Speech, 52*(4), 463–480. https://doi.org/10.1177/0023830909336584.
Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology, 102*(4), 522–537. https://doi.org/10.1016/j.jecp.2008.05.003.
Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Current Directions in Psychological Science, 25*(4), 291–296. https://doi.org/10.1177/0963721416655786.
Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio, 2*(2), 221–247. https://doi.org/10.1418/10879.
New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science, 19*(12), 1223–1227. https://doi.org/10.1111/j.1467-9280.2008.02228.x.
Pająk, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(9), 1377–1399. https://doi.org/10.1037/xlm0000247.
Pascalis, O., De Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296*(5571), 1321–1323. https://doi.org/10.1126/science.1070223.
Pater, J., Stager, C. L., & Werker, J. F. (2004). The perceptual acquisition of phonological contrasts. *Language, 80*(3), 384–402. https://doi.org/10.1353/lan.2004.0141.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies Retrieved from. *Spatial Vision, 10*(4), 437–442 https://www.denispelli.com/pubs/pelli1997videotoolbox.pdf.
R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.R-project.org
Redford, M. A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America, 106*(3, Pt 1), 1555–1565. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10489711
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.

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.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development, 49*(4), 1114–1128. https://doi.org/10.2307/1128751.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382. https://doi.org/10.1038/41102.

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental Psychology, 52*(7), 1011–1023. https://doi.org/10.1037/dev0000114.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166. https://doi.org/10.1016/s0010-0277(00)00081-0.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13*(5), 480–484. https://doi.org/10.1111/1467-9280.00485.

Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology, 153*, 140–148. https://doi.org/10.1016/j.jecp.2016.07.014.

Treiman, R., Broderick, V., Tincoff, R., & Rodriguez, K. (1998). Children's phonological awareness: Confusions between phonemes that differ only in voicing. *Journal of Experimental Child Psychology, 68*, 3–21. https://doi.org/10.1006/jecp.1997.2410.

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 Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition, 24*(5), 573–583. https://doi.org/10.3758/BF03201084.

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 246–254. https://doi.org/10.1037/a0025260.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234. https://doi.org/10.1207/s15473341lld0102_4.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*(1), 1–30. https://doi.org/10.1207/15250000252828226.

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.

White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language, 68*(4), 362–378. https://doi.org/10.1016/j.jml.2013.01.003.

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer Verlag.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12*(3), 412–418. https://doi.org/10.1111/j.1467-7687.2008.00789.x.