Two-Year-Olds Interpret Novel Phonological Neighbors as Familiar Words

Daniel Swingley
University of Pennsylvania

When children hear a novel word in a context presenting a novel object and a familiar one, they usually assume that the novel word refers to the novel object. In a series of experiments, we tested whether this behavior would be found when 2-year-olds interpreted novel words that differed phonologically from familiar words in only 1 sound, either a vowel or consonant. Under these conditions children almost always chose the familiar object, though examination of eye movements showed that children did detect the tested phonological distinctions. Thus, children discounted perceptible phonological variations when doing so permitted a resolution of the speaker's meaning without postulating a new word. Children with larger vocabularies made novel-word interpretations more often than children with smaller vocabularies did. The results suggest that although young children do interpret speech in terms of a learned phonological system, this does not mean that children assume that phonological distinctions imply lexical distinctions.

Keywords: language acquisition, lexical development, word recognition, phonology, speech perception

Conventional descriptions of language hold that all languages have a system for converting the speaker's physical movements of pronunciation, which are fundamentally "analog" or gradient, into abstract linguistic units, which are fundamentally "digital" or discrete. These discrete units, including consonants, vowels, tones, and stress patterns, specify and differentiate words, according to the phonology of the language. The physical movements from which these units are recovered also serve to signal nonlexical information relevant to the discourse. This information may be categorical (such as talker identity) or continuous (emotional state, degree of emphasis, speaking rate).

Successful interpretation of any utterance demands recovering the utterance's phonological description in terms of the discrete units that the talker intended to convey, and also placing the utterance into the relatively continuous landscape of discourse meaning. One might imagine that listeners accomplish this by identifying consonants and vowels, assembling them into words as they are decoded, and then interpreting the leftover variation to characterize talker identity and discourse structure. However, this cannot be the general case. During speech comprehension phonological descriptions cannot always, or even usually, be determined with certainty from the signal alone, and in response, listeners simultaneously make use of all the information at their disposal to arrive at a best guess at the speaker's meaning. The path from the canonical realization of a word (which comes from its

phonological description, on most accounts) to the actual auditory experience of the listener is shaped by phonetic context, sentence-level prosody, dialectal variation, ambient noise, and diverse articulatory shortcuts undertaken by the speaker for his or her own ease (e.g., Hawkins, 2010). But despite the unkempt reality of ordinary speech, the phonological idealization remains an important point of reference. If I am sure that you intended to say /parət/ ("parrot"), perhaps because of your clear, careful articulation, I can rule out the possibility that you might be talking about carrots or ferrets; these words are similar to "parrot," but cannot be "parrot" because the phonology specifies /p/, /k/, and /f/ as distinct. In this sense a "parrot" is no more a carrot than a "sneeze" is.

The present series of experiments concerns children's appreciation of this principle of phonological contrast, the notion that distinct sequences of phonological categories signify distinct meanings. The starting point for this work is that we should not take for granted that children understand phonological contrast per se simply by virtue of their perceptual abilities. The connection between sound categorization and meaning categorization is not something that is characteristic of perceptual interpretation in general; it is a peculiar property of language. Outside of language, for many things in the world, similar objects have similar functional relevance. The similarity of aspens and poplars, cinnamon and cassia, or hawks and falcons matches our general indifference to the distinctions between them. But because the connection between sound forms and meanings is almost entirely arbitrary, small changes (like the tiny acoustic difference between "carrot" and "parrot") can have a disproportionate impact on the function of the utterance.

How children arrive at this perspective on the relationship between phonological and semantic variation is not known. Research with infants shows that they learn to categorize and differentiate the speech sounds of their native language over the course of the first year of life (e.g., Kuhl, 2004; Werker, Yeung, & Yoshida, 2012), refining these abilities into the early school years (e.g., Hazan & Barrett, 2000; Medina, Hoonhorst, Bogliotti, & Serniclaes, 2010). One-year-olds can be taught pairs of similar-

sounding words that differ phonologically (Fennell & Waxman, 2011; Yoshida et al., 2009) but not words that are phonetically distinct but phonologically the same (Dietrich, Swingley, & Werker, 2007). And in general, phonological mispronunciations of words make those words harder to recognize for infants and children (e.g., Bailey & Plunkett, 2002; Swingley, 2005, 2009; Vihman, Nakai, DePaolis, & Hallé, 2004), but no such effect is found for equivalent phonetic changes without phonological significance (Ramon-Casas, Swingley, Bosch, & Sebastián-Gallés, 2009; Quam & Swingley, 2010). These findings all comport with the common view that once infants learn their language's phonological categories, these categories are, as a matter of course, active in the learning, representation, and recognition of words, and are available for defining lexical contrast.

But in fact these results do not require that children have mature intuitions about phonological contrast. They leave open the possibility that infants' learned phonetic categories provide a language-appropriate similarity space that they use to compare heard realizations of words with their expected, canonical phonetic forms—and that is all. A pronunciation like *vaby* might be a poor realization of *baby* without being an illegal one. For a mature speaker of English, the situation is different. *Vaby* heard where *baby* might have been expected demands explanation: Did the talker misspeak? Is it a joke or game? Did I misperceive the speaker? If not, then *vaby*, being a phonologically impermissible pronunciation of *baby*, is another word entirely, just as much as it is when appearing as a printed word, even if its proximity to *baby* reminds the listener of babies (Marslen-Wilson, Moss, & van Halen, 1996).

Another way to pose this question is to ask what is happening in children's minds when they perform in eyetracking tests of their sensitivity to mispronunciations—the empirical method that has been most commonly used to explore toddlers' interpretation of phonological variation in words. In a typical experiment, children see a pair of pictures on a display, and their eye movements are monitored as they hear sentences naming one of the pictures using a correct pronunciation ("Where's the baby?") or a deviant one ("Where's the vaby?"). Children fixate the named picture less when it is given a deviant pronunciation (e.g., Swingley & Aslin, 2000). Indeed, children fixating the target picture are more likely to defect from the target upon hearing a deviant pronunciation than upon hearing a canonical pronunciation, as if they were actively rejecting the label as a possible name for the target, and seeking a more suitable referent elsewhere (Swingley, 2009).

However, an alternative interpretation is that children's target-rejections are not motivated by an active search at all. Children's baseline behavior in this task is to look back and forth—this is what they do before the speech sample begins, and it is what they do if they understand nothing at all of the speech stimulus (Swingley & Fernald, 2002). To some degree, this gaze-alternation behavior is modified (toward fixating the target) to the extent that the spoken word matches the phonetics of a word the child knows. Shifting away from the target might not be motivated by a search for a novel object that could bear a name like *vaby*; it might simply be a return to children's baseline behavior. (When the spoken word does not sound like any other words, children do seek a novel object, but this does not require any phonological sophistication; Halberda, 2003).

Habituation studies of word learning can also be seen as demonstrations of a learned similarity space rather than manifestations of a phonological rule about contrast. Children habituated to an audiovisual stimulus in which an object image and a syllable repeatedly co-occur recover from habituation when the syllable is phonologically altered (e.g., from *bin* to *din*; Werker, Fennell, Corcoran, & Stager, 2002). But this behavior need not have as its foundation any assumption that phonological distinctions imply lexical differences. The task specifies this fact for the children; indeed, in 9- and 10-month-olds, this task environment can be powerful enough to induce lexical distinctions between words that do not contrast in the native language (Yeung & Nazzi, 2014; Yeung & Werker, 2009).

The approach taken in the experiments reported here was to examine children's spontaneous interpretation of phonological variants ("mispronunciations") of familiar words. Crucially, children were not instructed or shown that two phonologically distinct, but similar, words were contrastive; they were instead given the opportunity to reveal this interpretation themselves, through selection of a novel object rather than a familiar object as the referent of a novel word. For example, children were shown an image of a dog and a presumably unfamiliar object, and asked to select the *tog*. Children choosing the novel object would reveal their rejection of *tog* as a legitimate word for a dog. The method was based on a well-known, readily replicable behavior commonly referred to as "mutual exclusivity:" rejection of the use of two distinct words to refer to the same object, as manifested by the child's selection of a proffered novel object rather than a familiar one upon hearing a novel word (e.g., Markman & Wachtel, 1988). A virtue of this procedure is that the children's task is a straightforward one: We ask for an object, and they must choose which of two options is the correct referent. In our implementation of the task, children's eye movements were also monitored, to provide a measure of their online interpretation of the utterances (see also Creel, 2012).

Similar tests of mutual exclusivity under various levels of phonological similarity have produced somewhat mixed results, though greater deviation from the familiar word's phonology typically results in less identification of that form with the familiar object. One study employing mostly multiple-feature variants of familiar words found 4-year-olds interpreting such variants as novel names for unfamiliar objects nearly every time, and 2-year-olds doing so most of the time (Jarvis, Merriman, Barnett, Hanba, & Van Haitsma, 2004).

On the other hand, Merriman and Schuster (1991) found that 25-month-olds chose a familiar object rather than a novel object most of the time if the phonological "near miss" differed from the familiar form by a change of one phoneme, though results varied substantially from item to item. Similarly, White and Morgan (2008) found that 19-month-olds fixated a familiar object more than a novel object upon hearing a deviant (consonantal) pronunciation of a familiar word. In that study, consonants of different degrees of deviance (in number of phonological features) led to predictable declines in the amount children fixated the target, up to a three-feature difference for which children actually shifted to gazing more at the novel object. Creel (2012), using relatively subtle phonemic changes (such as [æ] for [ɛ] or [p] for [b]), found that preschoolers (3- to 5-year-olds) chose the familiar object more than 80% of the time.

Thus, children's interpretation of phonological "near misses" has a gradient character, and in spite of the facility with which infants and toddlers discriminate speech sounds (Werker, Yeung,

& Yoshida, 2012) and detect mispronunciations in word recognition tasks (Swingley, 2003), there is little evidence to suggest that children apply a strict phonological criterion in identifying which words are familiar and which are novel.

Here, we examined an intermediate age group, 2.5-year-olds, to test their tendency to spontaneously interpret perceptible phonological variants as the familiar word, or as a novel word. We tested variants generated by modifying either vowels or consonants. In the first experiment, we replicated the finding that very young children generally interpret close novel neighbors as their familiar counterparts even when a novel object is available (Swingley & Aslin, 2007; White & Morgan, 2008), and Creel's (2012) finding (in older children) that both fixation responses and manual picture-selection responses yield similar conclusions. In two additional experiments, we tested whether children could be induced to treat phonological variants as lexically contrastive, first with a pragmatic manipulation and then with a direct teaching manipulation.

We tested 2.5-year-olds because children this age are generally viewed as being competent perceivers of phonological distinctions (though this is infrequently tested, as this age group is less often evaluated in research studies than 1-year-olds or preschoolers are). In addition, younger children (1.5-year-olds) do not treat a novel phonological neighbor of a familiar word as lexically novel even when given apparently unassailable evidence supporting this interpretation (Swingley & Aslin, 2007; though see Dautriche, Swingley, & Christophe, 2015). Creel (2012) tested substantially older children, in a potentially more complex task involving four pictured alternatives. It is possible that with many alternatives, children would be more likely than otherwise to latch on to the referent whose label the spoken variant resembles, because of the relatively greater uncertainty that comes with a more complex response set. Here we employed two pictures to keep the task simple (as befits the young age group) and similar to prior work testing fixation responses to phonological variants.

Deviant pronunciations were created by starting from a set of familiar object labels, and modifying the first consonant (to make a consonant mispronunciation, c-MP) or the first vowel (to make a vowel mispronunciation, v-MP) of each word. In some studies, somewhat younger children, as well as adults, have appeared to interpret or recall consonants and vowels differently (Creel, Aslin, & Tanenhaus, 2006; Cutler, Sebastián-Gallés, Soler-Vilageliu, & Van Ooijen, 2000; Hochmann, Benavides-Varela, Nespor, & Mehler, 2011; Nazzi, 2005; Nazzi, Floccia, Moquet, & Butler, 2009; Nazzi & New, 2007). In others, consonants and vowels are treated similarly (Floccia, Nazzi, Delle Luche, Poltrock, & Goslin, 2014; Mani & Plunkett, 2007; Swingley & Aslin, 2002). Some of the discrepancy among these results may be due to variation in the methods, whereas some variation is systematically linked to the child's language (Højen & Nazzi, 2015). When a difference is found, it is generally in the direction of consonants seeming to exert a greater influence on children's explicit lexical decision-making than vowels, but neither being clearly more important than the other in word-recognition tasks. If this is true, children in the present experiments might show equivalent performance on consonant and vowel mispronunciations for the visual fixation measures but not the manual referent-selection responses.

Experiment 1

Method

Participants.   In this and all other experiments, children were recruited primarily via mailed invitations to parents, and also by meeting with parents in doctors' waiting rooms. Children were excluded if their parent reported the child hearing less than 75% English during a typical week. Parents all reported raising their children monolingually in English. Twenty-six children (10 girls) aged between 24 and 30 months (mean, 825 days; $SD = 57$) were retained in the final sample. An additional nine were tested but were excluded for being fussy during the session (i.e., crying, refusing to touch the screen, continually leaping off the parent's lap). Parents were given a vocabulary questionnaire to complete before their session (the Macarthur-Bates Communicative Development Inventory, Words & Sentences or CDI; Fenson et al., 1994). In each of the experiments reported here, about 34% of participants' parents identified their child as Black/African American, 57% White, and the remainder Asian, Hispanic, or Native American.

Visual stimuli.   The visual stimuli were photographs of objects on a gray background, presented side by side on a 43.2 cm touchscreen. Pictures were of similar sizes, averaging about 6.3 cm wide. Participants were seated about 45 cm from the screen, close enough that they could touch the screen with their fingers while remaining seated on their parent's lap. A set of short animations was created to reinforce correct touches to the screen during training and filler trials. These included such events as a monkey bouncing on a beach ball, a cat scooting by on roller skates, and a frog bounding over a log.

Auditory stimuli.   The speech stimuli were digitally recorded by a female native speaker of American English. Her speaking rate was slow and in an "infant-directed" register. Target words always appeared at the end of the sentence. Sentences included *Touch the* (target), *Where is the* (target)?, *Choose the* (target), *Which one is the* (target)?, and *Point to the* (target). A given target word and its variants always appeared in the same carrier phrase. Sentences were not spliced, but recorded individually, though different realizations of a given carrier phrase were very similar. The five target words' correct pronunciations and their consonant-altered and vowel-altered variants were: *apple*, *ackle* ([ækl]), *epple* ([ɛpl]); *book, pook* ([pʊk]), *buwk* ([buk]); *duck, guck* ([ɡʌk]), *dock* ([dak]); *fish, vish* ([vɪʃ]), *fiesh* ([fiʃ]); *kitty, pity* ([pɪɾi]), *ketty* ([kɛɾi]). Fillers and training words were *baby, ball, car*, and *dog*; non-neighbor nonce words were *meb* ([mɛb]), *shang* ([ʃæŋ]), *chome*, ([tʃom]) and *riz* ([ɹɪz]). Consonantal mispronunciations involved a single feature change and vowel mispronunciations involved vowels that are adjacent in phonetic space.

The familiar stimulus words were chosen because they are known to children from a young age. Much younger children (14–24 months) recruited in similar ways recognize these words in language-guided looking tasks (Romeo & Swingley, 2015; Swingley, 2009; Swingley & Aslin, 2000; Swingley, Pinto, & Fernald, 1999). Data from the CDI checklists collected for this study indicated that most of the children were saying most of the words (over the five words, 89.6%, likely an underestimate of their receptive knowledge of these words).

Children's word learning and word recognition are affected by the frequency of words' component sounds and how commonly

these sounds occur together, a property known as *phonotactic probability* (e.g., MacRoy-Higgins, Shafer, Schwartz, & Marton, 2014; Storkel, 2001). This property can be estimated several ways. The measure reported here is the mean summed log probability of each of the bigrams in each word, by word position, over the words in the Brent corpus of infant-directed speech (Brent & Siskind, 2001; Swingley & Aslin, 2007). For example, for the stimulus *vis*, the bigram /vɪ/ in word-initial position appears in *video, visit, visual* . . ., words whose log frequency in the Brent corpus was calculated and summed. For each bigram in (e.g.) *vis*, this figure was divided by the total log frequency of all words long enough to have a bigram in that position. The average of these bigram probabilities, for each bigram in the word, was taken as the measure of that word's phonotactic probability. By this metric, the experiment's stimulus words did not differ significantly by condition, $F(2, 12) = 1.04$, but phonotactic probability is included as a predictor in the item-level analyses presented after the final experiment.

Apparatus and procedure. Visual materials were shown on a touchscreen. The screen was mounted on a frame above a video-camera that was used for recording children's eye movements. The camera was concealed by a black cloth-covered shield. The parent held the child on her lap facing the screen. The parent's view of the screen was blocked by an opaque visor. The experimenter sat opposite the child behind the back of the touchscreen, controlling the experiment using a laptop computer. Speech stimuli were presented from two small loudspeakers placed to the left and right of the touchscreen.

The experiment consisted of 32 trials. Each trial began with the simultaneous presentation of two pictures on the screen. About 750 ms later, the speech stimulus began. When the sentence was complete, the pictures remained on the screen until the child touched one of the images.

The first four trials, and four additional trials dispersed throughout the trial orders, were training trials intended to teach and maintain the behavior of touching the named object. On these training trials, if children touched the named object, a short reward animation was played; if they touched the distracter object, the screen went dark and children were played a one-sentence recording saying one of a set of sentences like "Listen to the words" or "No, not that one" and the trial was repeated. During the initial trials children were verbally encouraged to touch the named target, and if necessary the experimenter came around to the front side of the touchscreen and modeled the target-touching behavior.

On the remainder of trials, the same reward animation was played regardless of which image children touched, and trials were not repeated based on children's behavior. Thus, children were reinforced for touching the target only on the filler trials, and were reinforced for all touches on the crucial test trials.

Four stimulus orders were created. Among test trials each familiar word was heard twice with its normal pronunciation and once with each of its deviant pronunciations (altered vowel or altered consonant); thus, there were a total of 10 correct-pronunciation test trials and 10 mispronunciation trials, plus four nonce trials (*meb, chome,* etc.), and eight filler trials including the four training trials that started the experiment. No word (or variant, e.g., *fish, vish*) occurred on consecutive trials; no images were repeated on consecutive trials; each image appeared on the left and right equally often; the left and right sides served as target equally

often in each condition; and the first and second halves of the experiment were populated similarly by each item and condition. The second order was the same as the first but with the consonant-altered soundfiles traded with the vowel-altered ones. The third and fourth orders were the same as the first and second but with the items rotated (e.g., book for duck, duck for kitty . . .), thereby counterbalancing the orders of the items relative to the other orders.

Coding. Videotapes of the participants' faces were stamped with a digital timecode and digitized to Quicktime format. Several highly trained coders used George Hollich's *SuperCoder* software (Hollich, 2008) to step through each video recording frame by frame, noting for each frame whether the participant was looking at the left picture, the right picture, or neither (i.e., in transit between pictures, or fixating off-screen). This response information was integrated with trial-timing information generated by tone pulses aligned with stimulus events, yielding an accurate record of the timing of responses to the target words.

Trials on which children were not fixating the screen for at least 20 frames in the first 2 s after target-word onset were discarded, which removed 4.2% of the trials. On 23 trials (3.7%) children did not give interpretable touching responses.

Results

The task given to the children can only illuminate their phonological interpretation if children recognize the words they are tested on when correctly pronounced, and if they interpret phonologically distant nonce words as referring to the novel object they are shown. To evaluate these preconditions, children's looking and touching behavior were examined. Looking time proportions were computed over the time window from 367 ms to 2,000 ms from the onset of the spoken target words, following prior standards (e.g., Swingley & Fernald, 2002). Manual responses were registered automatically by the touchscreen and verified from videotape. Simultaneous touches of both pictures were treated as missing data; sequential touches were counted as selections of the first picture touched. Touching proportions were computed out of touches to either object. Looking proportions were computed by summing looking time to the target and dividing this by the looking time to the target plus looking time to the distracter.

Both of the experiment's preconditions were met, both in terms of children's fixation proportions, and their manual responses. The mean proportion of fixation to the target pictures, by subjects, on correct-pronunciation (CP) trials was 73.5%, significantly greater than 50%, $t(25) = 9.39$, $p < .0001$, 95% CI [69.2, 100] with 24 of 26 children and five of five items above 50% (min. item 65.6%). The mean proportion of correct touches to the named target on CP trials was 83.0%, $t(25) = 7.91$, $p < .0001$, 95% CI [75.8, 100], with 21 of 26 children and five of five items above 50% (min. item 76.0%). Thus, on the whole, when children heard a correct pronunciation of a familiar word, they looked at and touched the corresponding familiar object (see Figure 1, left side, for looking data; Figure 2 for touching data).

When children heard a nonce word like "chome" they looked at the displayed familiar object less than 50% of the time (mean, 39.2%; $t(25) = 3.16$, $p < .0025$, 95% CI [0, 45.0]), with 21 of 26 children below 50%. Children touched the familiar object significantly less than half of the time (mean, 27.9%, $t(25) = 3.74$, $p <$
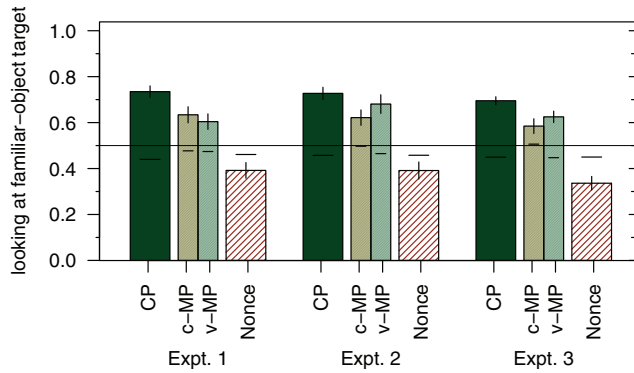
Figure 1. Target-looking proportions in all three experiments. The *y*-axis gives the average proportion of time children fixated the named target (CP trials, dark bars), or the familiar object (MP trials: lightly shaded bars; nonce trials: striped bars). The MP trials are divided into those with altered consonants (left shaded bar) and those with altered vowels (right shaded bar). Error bars show the standard error of the mean. For each displayed condition, a small horizontal line marks the mean target fixation in the early portion of the trial before the target word was spoken. CP = correct-pronunciation; c-MP = consonant mispronunciation; v-MP = vowel mispronunciation; Expt. = Experiment. See the online article for the color version of this figure.

.0005), with 21 children below 50% (and three at 50%). In sum, when children heard a nonce word not resembling any familiar words, they apparently interpreted it as referring to the unfamiliar displayed object, which they looked at and touched.

When children heard a mispronounced (MP) variant of a familiar word, they generally did not treat it as a novel word referring to the nonce object. Familiar-object fixation averaged 63.4% (consonant MPs) and 60.4% (vowel MPs), both significantly above 50% (c-MP $t(25) = 3.83$, $p < .0005$, 95% CI [57.4, 100]; v-MP $t(25) = 3.08$, $p < .0025$, 95% CI [54.6, 100]). Children's manual selections showed a similar pattern. Familiar-object touching averaged 72.3% (c-MPs) and 72.1% (v-MPs), both significantly above 50% (c-MP $t(25) = 5.02$, $p < .0001$, 95% CI [63.1, 81.4]; v-MP $t(25) = 3.92$, $p < .0005$, 95% CI [60.5, 83.6]).

Although children most often fixated the named (familiar) object on MP trials, the decrement in target looking due to the mispronunciation was significant (CP vs. c-MP, mean difference 10.1% ($SD = 18.5$), $t(25) = 2.77$, $p = .010$, 95% CI [2.6, 17.6]); CP versus v-MP, mean difference 13.1% ($SD = 13.1$), $t(25) = 3.52$, $p < .002$, 95% CI [5.4, 20.7]). Target fixation given consonant mispronunciations was similar to target fixation given vowel mispronunciations (c-MP vs. v-MP, mean difference 3.0% favoring c-MP ($SD = 21.1$), $t(25) = 0.72$, *ns*). Children also touched the target image less when they heard either sort of mispronunciation. The difference for consonant-changed variants was 10.6% ($SD = 24.7$), $t(25) = 2.20$, $p < .037$, 95% CI [0.7, 20.6] and for vowel-changed variants 10.9% ($SD = 23.4$), $t(25) = 2.38$, $p < .025$, CI [1.5, 20.3].

Effects of children's age and CDI (vocabulary) counts are presented in an omnibus cross-experiment analysis after discussion of Experiment 3.

These results show that even when children's word recognition systems show sufficient sensitivity to a mispronunciation to divert attention away from the word's referent, this sensitivity does not often trigger the child to settle on a novel-word interpretation of what he or she has heard.

One way around this conclusion is to suppose that there are two sorts of MP trials: those on which children entirely fail to detect the MP (and therefore show no looking or touching consequences of it), and those on which children detect the MP and make the linguistic inference that a novel word was spoken. This account makes the prediction that familiar-object touches on MP trials come mainly from the trials on which children failed to detect the MP. But this is false: considering only trials on which children chose the familiar object, there was still a significant effect of mispronunciation on fixation to the target (mean CP, 79.6%; mean c-MP, 73.3%; mean v-MP, 71.2%; both paired differences from CP significant, $t(24) > 2$, $p < .03$). Thus, on many trials, the "error signal" provoked by the MP was sufficient to temporarily disrupt the connection between the spoken word and the word's meaning, but this error signal was overridden by children in choosing the familiar object, rather than being interpreted as a sign that a new word was offered.

In sum, Experiment 1 provided very little indication that children consider minor phonological deviations in familiar words to constitute novel lexical items for which novel-object interpretations are appropriate. This result replicates some of the findings of White and Morgan (2008) and Creel (2012), in children of an intermediate age and using a two-choice procedure integrating looking and touch responses.

In Experiment 2, we consider the possibility that children might be encouraged to adopt novel-word interpretations if they were provided evidence that the speaker had superior familiarity with the words she was using and with the objects she was referring to. This manipulation was intended to address the possibility that children typically deploy a phonological-difference criterion when determining which forms are novel words, but did not do so in Experiment 1 because they did not have any basis for believing the speaker to be familiar with the novel objects.
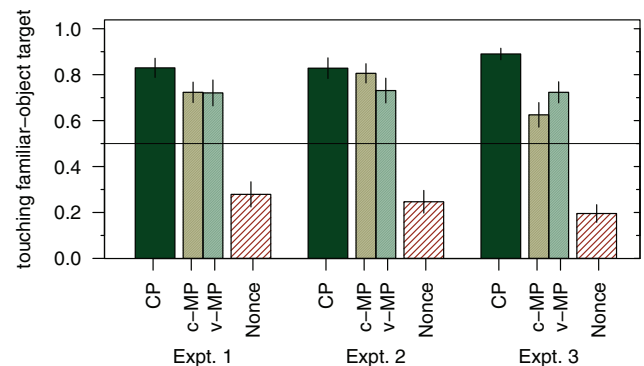

Figure 2. Target touching proportions in all three experiments. The *y*-axis gives the average proportion of time children manually selected the named target (CP trials, dark bars), or the familiar object (MP trials: shaded bars; nonce trials: striped bars). The MP trials are divided into those with altered consonants (left shaded bar) and those with altered vowels (right shaded bar). Error bars show standard error of the mean. CP = correct-pronunciation; c-MP = consonant mispronunciation; v-MP = vowel mispronunciation; Expt. = Experiment. See the online article for the color version of this figure.

Experiment 2

Two features distinguished this experiment from the preceding one. First, in an introductory phase, children were shown a cartoonish drawing of a woman and heard, "Welcome, boys and girls. We're going to play a game today. I have lots of fun things to play with. I have these toys." At this point five of the novel object images appeared on the screen. "And I have these toys." Five more of the novel objects appeared. "And I play with these toys." Then the remaining five novel objects appeared. "I got them at a really special toy store." Finally the familiar objects appeared: "Here are some more. I have all of these things in my house and I play with them all the time." The objects were taken away, leaving only the image of the woman: "Are you ready to play the game?"

The other new feature of the second experiment was series of statements played before each test trial, always using the same woman's voice: "I know the words for both of these things. I have these toys in my house. I know what both of these things are called."

These manipulations were intended to provide evidence to the children that the speaker possessed privileged knowledge of the novel toys, including knowledge of how the toys should be called. We supposed that this might give children greater confidence that the novel words employed by the talker were legitimate reflections of her superior knowledge. Related manipulations have been shown to affect children's interpretation of potentially ambiguous language (e.g., Henderson, Sabbagh, & Woodward, 2013; Krogh-Jespersen & Echols, 2012; review, Harris & Corriveau, 2011; Sobel & Kushnir, 2013). For example, 18-month-olds (Brooker & Poulin-Dubois, 2013) and 24-month-olds (Koenig & Woodward, 2010) are less likely to learn a novel word if it was taught by a talker who had previously spoken inaccurately about the names for objects. Of course, distrusting inaccurate speakers more than neutral ones is not the same as putting more faith in especially knowledgeable ones, so whether a similar effect would be shown here was not known.

In all other respects the conduct of Experiment 2 was identical to that of Experiment 1.

Participants

Twenty-six children (12 girls) aged between 24 and 30 months (mean, 845 days; $SD = 48$) were retained in the final sample. Recruitment was as in Experiment 1. An additional 18 children were tested but excluded from the sample for being fussy (10), parental or sibling interference (four), or technical problems with the recordings (four). We failed to obtain a completed CDI from one parent. 13.9% of trials were removed for insufficient looking (see Experiment 1) and children did not give interpretable touching responses on 37 trials (6.9%).

Results

As in Experiment 1, children's proportion of fixation to the target was well above 50% on CP (correct pronunciation) trials (mean, 72.7%, $SD = 13.6$, $t(25) = 8.54$, $p < .0001$, 95% CI [68.1, 100]), with 23 of 26 children and five of five items above 50% (min. item 64.3%). The mean proportion of touches to the target on CP trials was 82.8% ($SD = 22.9$, $t(25) = 7.91$, $p < .0001$, 95% CI [75.8, 100]), with 24 of 26 and five of five items above 50% (min. item 72.0%). Children's proportion of fixation to the familiar object on nonce trials was 39.1% ($SD = 19.0$), significantly less than 50%, $t(25) = 2.92$, $p < .004$, 95% CI [0, 45.4], and they touched the familiar object less than half of the time (mean 24.7%, $SD = 24.5$, $t(25) = 3.73$, $p < .0005$, 95% CI [0, 38.0]), with 19 children below 50% (and three at 50%). Thus, overall, when children heard a familiar word, they looked at and touched the named picture, and when they heard a nonword not resembling a familiar word, they looked at and touched the novel object picture (Figures 1 and 2, middle panels).

When children heard mispronounced (MP) versions of the target words, target (familiar-object) fixation was well above 50%: for consonant MPs, 61.4% ($SD = 18.0$), $t(25) = 3.25$, $p < .002$, 95% CI [55.4, 100]); for vowel MPs, 68.7% ($SD = 19.7$), $t(25) = 4.85$, $p < .0001$, 95% CI [62.1, 100]). Manual selections were also above 50%: c-MP mean 78.6% ($SD = 21.7$), $t(25) = 5.02$, $p < .0001$, 95% CI [64.7, 100]); V-MP mean 75.0% ($SD = 27.5$), $t(25) = 3.92$, $p < .0005$, 95% CI [62.4, 100]).

As in Experiment 1, target fixation was greater on CP than MP trials. This effect of mispronunciation was significant for the c-MP condition (mean effect, 11.3%, $SD = 20.0$, $t(25) = 2.88$, $p < .01$, 95% CI [3.2, 19.3]), but not the v-MP condition (mean effect, 4.0%, $SD = 18.9$, $t(25) = 1.07$, ns). The difference between these effects (7.3%, $SD = 25.3$) was not significant, $t(25) = 1.47$, $p = .15$, 95% CI [−2.9, 17.5]. The overall effect of mispronunciation was significant if the c-MP and v-MP subconditions are collapsed (mean 7.4%, $SD = 16.2$, 95% CI [1.4, 13.8]). Thus, though the evidence for fixation effects of consonant mispronunciations was stronger than for vowel mispronunciations, there was not a reliable basis for arguing that children were more sensitive to the consonant changes.

Children's touching of the familiar object was not significantly affected by mispronunciations, though familiar object selection was numerically inferior on MP trials. The mean difference for consonant changes was 4.2% ($SD = 30.3$, $t(25) = 0.7$, ns; 95% CI [−0.8, 16.4]) and for vowel changes, 7.8% ($SD = 24.8$, $t(25) = 1.6$, ns; 95% CI [−2.2, 17.8]).

In summary, the pattern of results in the second experiment replicated the most important features of the first experiment: hearing consonant and vowel substitutions in familiar words did not lead children to assume that they were hearing a new word that would likely indicate an alternative referent. They did show a "mutual exclusivity" response upon encountering phonologically very distinct words, but not neighbors of familiar words.

Experiment 3

In Experiment 3, we tested a much less subtle and more directive manipulation. Starting from the materials of Experiment 1, we added an introductory segment in which five of the 10 novel objects were explicitly labeled and explicitly contrasted with their familiar neighboring word. The referent object appeared alone on the screen, and the same talker who voiced the test trials said, for example, "This is called a guck. Not a duck. This is a guck." In this way, children were taught our word for an object with a name

similar to each of the five real words.[1] They were then tested, as in Experiment 1, on all 10 novel words and their familiar counterparts. If children believe that phonological differences are sufficient to contrast words, but hesitated in the prior experiments because of uncertainty about the speaker's meaning in this context, explicit declaration about contrast might be expected to lead children to a contrastive interpretation. On the other hand, if children are strongly resistant to learning close novel neighbors, even this explicit manipulation might fail to change their minds.

Method

Design. For half of the children, two of the explicitly taught words were the vowel-changed variant of a real word, and the remainder were the consonant-changed variant. For the other half, three of the taught words were the consonant-changed variant. Thus, half of the children were taught *epple, dock, ketty, pook,* and *vish*, and the other half were taught *ackle, guck, pity, buwk*, and *fiesh*. Except as noted above, method was as in Experiment 1.

Participants. Twenty-six children (13 girls) aged between 24 and 31 months (mean, 853 days; $SD = 57$) were retained in the final sample. An additional 13 children were tested but not included: 10 were fussy, and in three cases parents interfered by finding ways to answer for their children on some test trials; 2.4% of trials were removed for insufficient picture fixation, and on 10 trials (1.6%) children did not give interpretable touching responses.

Results

Children's proportion of fixation to the target was well above 50% on CP (correct pronunciation) trials (mean, 69.5%, $SD = 8.8$, $t(25) = 11.3$, $p < .0001$, 95% CI [66.6, 100]), with all 26 children and all five items above 50% (min. item 64.3). The mean proportion of touches to the target on CP trials was 89.0% ($SD = 12.7$, $t(25) = 15.7$, $p < .0001$, 95% CI [84.8, 100]), with all 26 children and all five items above 50%. Fixation to the familiar object on nonce trials was 33.6% ($SD = 15.1$), significantly less than 50%, $t(25) = 5.5$, $p < .0001$, 95% CI [0, 38.6], and touching of the familiar object on nonce trials was below 50% (mean 19.6%, $SD = 20.5$), significantly less than 50%, $t(25) = 7.6$, $p < .0001$, 95% CI [0, 26.4], with 20 children below 50% and the rest at 50%. Thus, once again children hearing familiar words looked at and touched the named object, and children hearing unknown words looked at and touched the novel object (Figures 1 and 2, right panels).

Mispronouncing consonants and vowels reduced both familiar-object looking, and touching, to a significant degree. Children looked at the familiar object 11.1% less ($SD = 15.9$) when hearing a word with a consonantal substitution (paired $t(25) = 3.54$, $p < .001$, 95% CI [5.7, 100]) and 7.0% less ($SD = 12.3$) than when hearing a word with a vowel substitution, $t(25) = 2.9$, $p < .005$, 95% CI [2.9, 100]. These effects were not significantly different from one another, $t(25) = 1.0$, $p = .33$, 95% CI [−12.2, 4.1]. Looking to the familiar object was above 50% for both types of substitution (consonants, 58.4%, $t(25) = 2.6$, $p < .008$, 95% CI [53.0, 100]; vowels, 62.5%, $t(25) = 5.0$, $p < .001$, 95% CI [58.2, 100]). Children's touching of the familiar object was also reduced by mispronunciations of both types: consonants, mean effect 26.5% ($SD = 22.2$, $t(25) = 6.1$, $p < .0001$, 95% CI [17.5, 35.5]);

vowels, mean effect 16.7%, $SD = 22.4$, $t(25) = 3.8$, $p < .001$, 95% CI [7.6, 25.8]). Children touched the familiar object less on c-MP trials (62.5%, $SD = 27.2$) than on v-MP trials (72.3%, $SD = 23.4$), a borderline-significant difference, $t(25) = 2.06$, $p = .0502$, 95% CI [0, 19.6]. Still, children touched the familiar object more than half of the time (consonant MPs, 62.5%; $t(25) = 2.3$, $p < .015$, 95% CI [53.4, 100]; vowel MPs, 72.3%, $t(25) = 4.87$, $p < .0001$, 95% CI [64.4, 100]).

In Experiment 3, half of the test words had been explicitly taught in the introductory phase. For each child, and each familiar word like *fish*, one of its variants (e.g., *vish*) had been taught, as in *This is called a vish. Not a fish. This is a vish.* The other variant (e.g., *fiesh*) had not. A paired comparison of familiar-object gaze across taught and untaught words showed that teaching a novel word at the start of the experiment did not make children significantly less likely to fixate or choose the familiar object upon hearing that novel word. The mean familiar-object fixation for taught nonce words was 57.6% ($SD = 14.6$) and for untaught nonce words 63.8% ($SD = 14.3$), a small difference in the expected direction, but not a significant one (paired $t(25) = 1.7$, $p = .102$, 95% CI [−13.8, 1.3]); only 14 out of 26 children looked more at the familiar object for untaught novel words than for taught ones. Fixation to the familiar object was significantly above 50% even for the taught novel words, $t(25) = 2.6$, $p < .008$, 95% CI [52.7, 100].

Turning to the effect of word-teaching on the touching results, we found that children touched the familiar object somewhat less often when the novel-object alternative had been taught (mean, 62.2%, $SD = 31.0$) than when it had not (mean, 71.1%, $SD = 23.9$); however, this difference was neither significant, $t(25) = 1.4$, $p = .162$, 95% CI [−23.6, 4.2], nor consistent across children. Ten showed this pattern, seven were in the reverse direction, and the remainder showed no difference.

To summarize, children's performance in Experiment 3 was broadly quite similar to their performance in the preceding experiments. For the most part, close phonological neighbors were usually treated as acceptable versions of their familiar source words.

Comparison of the Three Experiments

To quantify the effects, if any, of the manipulations that differentiated the three experiments, the three datasets were combined and evaluated together using experiment as a factor. The trial-by-trial looking-to-target proportions were not normally distributed (about a third of the proportions were zero or one, which is typical of language-guided looking results in toddlers), so looking on CP and MP trials was analyzed over Subject × Condition Means rather than over trials. Trial-by-trial object touching has a binary outcome (touching the target or not) and was evaluated using multilevel hierarchical logistic regression (R version 3.1.0, R Core Team, 2015; Kuznetsova, Brockhoff, & Christensen, 2014). Power

---

[1] This training was not intended to test whether it is *possible* for children to learn minimal pairs. It is (e.g., Dautriche et al., 2015; Fennell & Waxman, 2011), though for most children it would probably require more exposure. Rather, the training was intended to make it utterly incontestable to children—if they are receptive to such a message—that two different words might sound very similar.

analysis indicated an 80% probability of detecting ($p < .05$) an effect of size 0.38 or greater, which was deemed adequate given the large impact the manipulations were expected to have. Nonce trials were analyzed separately from the CP and MP trials because the CP and MP conditions were matched over items (e.g., *kitty, pitty, ketty* formed a set based on *kitty*) and the nonce items (*meb, shang, chome,* and *riz*) were not.

Target fixation on nonce trials did not differ significantly across the three experiments, whether considering means over subjects or over items. An ANCOVA over subject means including experiment and *z*-score of CDI total yielded no effect of experiment: $F(2, 71) = 1.36$, ns, and a significant effect of CDI total $F(1, 71) = 5.78$, $p < .02$, such that children with larger vocabularies fixated the familiar object less often than children with smaller vocabularies (simple correlation $r = -0.29$, $p = .012$). Over items, an ANCOVA including Experiment and phonotactic probability of the four nonce words as predictors yielded no effect of experiment ($F(2, 8) < 0.3$) and no phonotactic effect:, $F(1, 8) = 1.27$, ns.

Likewise, touching of the familiar object on nonce trials did not differ significantly across experiments. Logistic regression analyses of children's touching responses, here and for the CP/MP trials below, began by including as random effects subject and target-word (both with random slopes for each condition and experiment, when applicable), and then models were simplified by removing predictors when justified by model comparison using likelihood ratio tests. Considering the nonce trials, model comparison reduced the predictors to experiment (retained because it was the point of the analysis) and *z*-score of CDI. Experiment was not a significant predictor of object-touching (see Appendix A for details.) Children with larger vocabularies chose the familiar object less often on nonce trials than children with smaller vocabularies did (correlation $r = -0.290$, $t(73) = -2.58$, $p < .015$). Over 40% of children touched the novel object every time, and these children were quite evenly distributed across the experiments (10, 10, 12). Thus, the experimental manipulations had no significant effects on children's behavior on nonce trials.

Turning from the nonce trials to the CP and MP trials, we find that target fixation on CP and MP trials did not vary significantly by experiment. Hierarchical regression models were fit including experiment, condition, *z*-score of CDI vocabulary count, interactions between experiment and condition, and between condition and CDI. There were no significant effects of experiment (max $|t| = 1.1$, min $p > .25$). Mispronunciation had a significant effect on target looking, as expected, $t(71) = 4.49$, $p < .0001$. The effect of mispronunciation was greater for children with higher CDI totals, $t(71) = 2.3$, $p < .025$; the simple correlation between subjects' CDI score and familiar-object looking on MP trials was -0.269 ($p < .025$). No other effects were significant. The full model's results are given in Appendices B and C.[2]

Of the 26 children in each experiment, the number whose MP-trial behavior was closer to their CP looking (i.e., fixating the named familiar object) than it was to their nonce looking (i.e., fixating the novel object) was, for the three experiments, 18, 18, and 20, respectively. Thus, across experiments, in their looking patterns children treated the novel neighbor more like an instance of its familiar source word than they treated it like a nonword.

Considering the five target words across experiments, mean familiar-object fixation on CP trials was greater than it was on c-MP and v-MP trials for every word. A hierarchical regression

over item means by experiment, target word, and condition yielded no significant effects involving experiment. In this item analysis, children were significantly more likely to fixate the familiar target when the spoken word's phonotactic probability was higher, $t(18.7) = 2.56$, $p < .02$, and fixated the target word less when hearing a mispronunciation, c-MP: $t(32.4) = -1.95$, $p = .06$; v-MP: $t(33.6) = -2.07$, $p < .05$.

Thus, the experimental manipulations, namely the speaker of Experiment 2 describing her familiarity with the objects and their labels, and the explicit training of Experiment 3, did not have a reliable impact on children's looking behavior.

Analysis of children's manual object choices, however, revealed a stronger effect of mispronunciation on familiar-object touching in Experiment 3 than in Experiment 1. Altered pronunciation reduced familiar-object touching by 10.3% in Experiment 1, 5.9% in Experiment 2, and 21.7% in Experiment 3. Comparison of multilevel logistic regression models resulted in a model containing experiment and condition (and their interaction), CDI and its interaction with Condition, phonotactic probability, and random effects of subject and Word (see Appendix D). Performance on CP trials did not vary across experiments, but the effect of mispronunciation was stronger in Experiment 3 than in Experiment 1 ($z = -2.63$, $p < .01$). That said, in all three experiments most children's touching patterns on MP trials were closer to their CP trial performance than their nonce trial performance ($n = 22, 19$, and 19 of 26 children). CDI did not have an impact on CP trial performance, but on MP trials, children with higher CDI scores were less likely to touch the familiar object, making more novel-word attributions ($z = -2.68$, $p < .01$). Finally, children chose the familiar object more often when the spoken word's phonotactic probability was higher ($z = 2.08$, $p = .04$).

The relationship between spoken vocabulary size (CDI) and children's looking and touching behavior is shown in Figure 3.

To summarize, comparison of the three experiments indicated that the results were substantially similar over the manipulations. In particular, although children could hear the phonological difference between the canonically pronounced (CP) and variant (MP) words, they usually assumed that the speaker was referring to the familiar object whether hearing CP or MP words.

The modest but significant increase in novel-object selection produced by the explicit teaching manipulation of Experiment 3 showed that children were not wholly rigid in their interpretation and that familiar-object selection is not some kind of automatic bias immune to contextual features of the broader situation. Given that this increase was not limited to the specific words we taught each child, the effect was most likely due to a change in children's willingness to accept close novel neighbors in general, and not localized to the words or contrasts we trained. If this finding turns out to be generally true, it suggests that rather than refining children's interpretation of, for instance, the [f]–[v] difference, the manipulation tuned a more generic sensitivity. At this point we do not know which features of the training mattered, but it seems likely that the explicit contrast played a role ("this is called a *vish*;

---

[2] The results are essentially the same if c-MP and v-MP are distinguished as separate levels of condition: Both are significantly different from CP, and there are no effects of experiment. CDI was associated with less familiar-object looking on c-MP trials, $t(142) = -2.05$, $p < .05$ but not v-MP trials, $t(142) = -0.93$, $p = .17$.
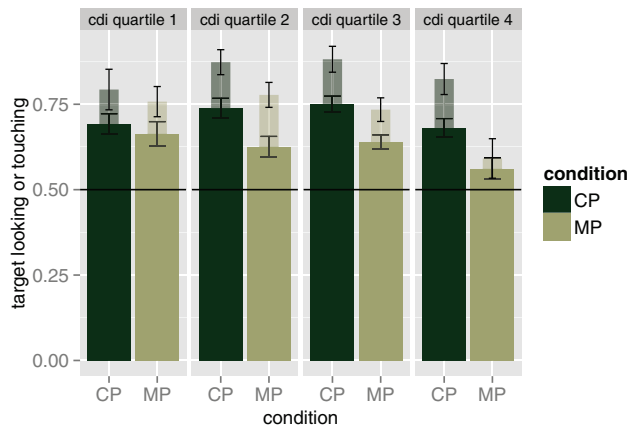
Figure 3. Target-looking and target-touching proportions in all three experiments, split by which quartile each child fell into in total vocabulary size (CDI). Wider bars indicate target looking proportions and narrow bars indicate target touching proportions. Dark bars show performance on correct-pronunciation trials (CP) and lighter bars show performance on mispronunciation trials (MP). Error bars show standard error of the mean by subjects. CP = correct-pronunciation; MP = mispronounced; cdi = Communicative Development Inventory. See the online article for the color version of this figure.

not a fish"). Even with this training, though, the dominant response was to treat novel words as if they were their more familiar competitors.

General Discussion

Phonological descriptions of languages include categories and rules that specify when two utterances contain the same set of words, or different sets of words. The English sounds /p/ and /t/ are said to *contrast* because the difference between them is sufficient for signaling a distinction between words. A *tail* is not a *pail*, and a novel syllable like *pree* cannot mean *tree* unless the speaker or listener has made an error.

In the present experiments, children of an age at which learning several words a day is normal detected small but phonologically distinctive phonetic deviations from familiar words, as shown by their reduced looking to, and less choosing of, the named pictures. Similar results have been found in children before (e.g., Swingley, 2009). However, children did not infer that the deviations could be resolved by assuming that they referred to unfamiliar objects. In this sense, children resembled adults who have low confidence in their memory representations of words they have heard only a few times (White, Yee, Blumstein, & Morgan, 2013).

On the other hand, children here were being tested on phonological deviants of words that we would expect to be among their "best," most familiar words, and the ones that they appear to have encoded faithfully. What do children think is happening when they hear mispronounced words in these studies? A natural supposition is that children would think that these forms are novel words. They sound different from the familiar words; in some cases a novel object is present; and children readily make the "mutual exclusivity" inference when hearing a word like *meb*. But this does not seem to be what children do, at least most of the time, when hearing novel phonological neighbors in this task.

Two interpretations of these findings are most plausible. The first is that children generally do not assume that words with slightly differing phonological descriptions are different words, and instead initially apply a more stringent phonetic difference criterion. Under this criterion, to be counted as a new word the novel form must be more different from a familiar word than one phonological feature. Indeed, it is possible that rather than having a phonological criterion, children use a gradient criterion in which the likelihood of a chunk of speech being a novel word rises as the phonetic distance between the new form and any existing forms increases (White & Morgan, 2008). If this is true, the receptive phonology of 2-year-olds does not match the conventional description of phonological systems as defining lexical contrast.

Such a gradient criterion could come about if children do not have well-defined phonetic categories in the first place. For example, if a language's consonant and vowel categories are not learned with much fidelity until children have large vocabularies (e.g., Edwards, Beckman, & Munson, 2004), they might show language-specific discrimination and categorization of clear exemplars (as in numerous infant speech perception studies; Werker, Yeung, & Yoshida, 2012) without parsing words into clear, well-defined analytic categories suitable for marking lexical distinctions.

A second possibility is that 2-year-olds do, at least sometimes, apply the traditional phonological difference rule, but not in the challenging context evaluated in the present study. Lexical activation from the image of the familiar word might overwhelm children's ability to consider a novel-word interpretation of the closely neighboring form. Or, more broadly speaking, children might consider, consciously or not, a wide range of features of the situation in estimating the probability that the form corresponds to a novel word—in particular, the likelihood that a novel object label would be introduced in the presence of a similar-sounding alternative (cf. Dautriche et al., 2015).

The manipulations of Experiments 2 and 3 were designed to revise children's estimates of this probability, by making the speaker a superior knowledge source (Experiment 2) or by showing, as explicitly as possible, that the novel words sounded like, but were not the same as, their familiar counterparts (Experiment 3). The general ineffectiveness of these attempts is not decisive against the position that children have a linguistically appropriate phonological-difference criterion that is masked by features of the situation, but it does challenge this position. Again, this is not to say that children cannot learn similar-sounding words; given sufficient evidence, they can, though most demonstrations of this capacity involve pairs of words that are both novel (e.g., Werker, Fennell, Corcoran, & Stager, 2002; Yoshida, Fennell, Swingley, & Werker, 2009). By 5- or 6-years-old, children can succeed in a task similar to that of Experiment 3 (Giezen, Escudero, & Baker, 2015). But the present experiments show that it is unlikely that small, phonologically relevant distinctions generally lead 2-year-old children to search for novel lexical items, certainly not in the eye-movement procedures that have been used most predominantly for evaluating children's receptive phonological knowledge.

How, then, do young children interpret the mispronunciations? They may shrug them off without drawing any conclusions. They may also consider them to be accented versions of words (Mulak, Best, Tyler, Kitamura, & Irwin, 2013; Schmale, Seidl, & Cristia, 2014). If so, it is surprising that children were, for the most part,

equally unlikely to treat vowel alterations and consonant alterations as lexically significant. In American English, the vowels bear much of the distinction among accents and among talkers (Labov, Ash, & Boberg, 2005), and if children note that the talkers in their environment produce the vowels of words differently, they might be inclined to discount vowel variation relative to consonant variation (though see Durrant, Delle Luche, Cattani, & Floccia, 2014). But such effects were minor here.

Discounting of variation in words' vowels might also be expected based on several studies showing that in some situations, adults rely more on consonants than vowels in identifying words (e.g., Cutler, Sebastián-Gallés, Soler-Vilageliu, & Van Ooijen, 2000). English-learning 30-month-olds showed a similar bias toward consonants in a word-matching task (Nazzi, Floccia, Moquet, & Butler, 2009). Children were given three words for three objects and asked to group the objects that belonged together, where the only consistent basis for judgment was word similarity. English-learning 30-month-olds, unlike younger English-learning children, tended to place together objects labeled by words varying in their vowels (*dib*—deb) rather than words varying in their consonants (*dib*—gib). This tendency reflects children's learning, because its developmental course varies from language to language (Højen & Nazzi, 2015). If children were learning that in English, consonants provide a stronger signal of lexical distinction than vowels, we would have expected more novel-object selection when our words varied from their familiar neighbor in the vowel rather than in the consonant. There were hints of this in Experiment 2 and a significant such effect on touching in Experiment 3, but not a consistent effect across the experiments. Perhaps children are beginning to learn this bias (which is present in direct conflict tasks, and in several tasks testing adults) but it is still weak at 2.5 years.

Children with larger vocabularies were somewhat more likely than children with smaller vocabularies to treat novel words as referring to novel objects, a finding similar to that of Law and Edwards (2015). The vocabulary effect on novel-object looking was as strong for nonce trials ($r = .29$) as for MP trials ($r = .27$), which might suggest that the vocabulary effect reflects a difference in ability or disposition to treat new words as words, and not primarily an effect of fine phonological interpretation per se. On the other hand, this disposition might itself come about more readily in children whose phonological interpretation skills are most secure because their categories are more robust or because they have a stronger insight into their function. There is now a substantial body of empirical work supporting links between vocabulary size and several lexical skills like word recognition and phonological processing in very young children (e.g., Law & Edwards, 2015; Marchman & Fernald, 2008, though vocabulary size is not always found to be linked to phonological sensitivity in word recognition (e.g., Swingley & Aslin, 2000). The causal pathways that yield these correlations, when found, are unclear at present.

The fact that a single phonological feature fails to render a word categorically distinct does not mean that children rely on a naive, language-general phonetic space when reasoning about lexical differences. The perceptual adaptation that infants undergo renders native-language distinctions more prominent than some non-native distinctions, and this learning process has direct consequences for word recognition and word learning (e.g., Dietrich et al., 2007; Quam & Swingley, 2010; Ramon-Casas et al., 2009). Learning

does guide interpretation. But the learning process is not complete at 2.5 years. The adapted perceptual space appropriate for the native language may not itself provide rules for its use in connecting speech to lexical differences. The determinants of this interpretive process, which may not be the same as the determinants of successful perceptual categorization of speech sounds, are not yet known.

References

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17,* 1265–1282. http://dx.doi.org/10.1016/s0885-2014(02)00116-8

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* B33–B44. http://dx.doi.org/10.1016/s0010-0277(01)00122-6

Brooker, I., & Poulin-Dubois, D. (2013). Is bird an apple? The effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy, 18,* E46–E68. http://dx.doi.org/10.1111/infa.12027

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

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. http://dx.doi.org/10.1016/j.jml.2005.09.003

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition, 28,* 746–755. http://dx.doi.org/10.3758/bf03198409

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel neighbors: Syntactic category matters. *Cognition, 143,* 77–86. http://dx.doi.org/10.1016/j.cognition.2015.06.003

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 16027–16031. http://dx.doi.org/10.1073/pnas.0705270104

Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2014). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language, 42,* 447–465. http://dx.doi.org/10.1017/S0305000914000063

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47,* 421–436. http://dx.doi.org/10.1044/10924388.2004.034

Fennell, C. T., & Waxman, S. R. (2011). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development, 81,* 1376–1383. http://dx.doi.org/10.1111/j.1467-8624.2010.01479.x

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59,* Serial Number 242. http://dx.doi.org/10.2307/1166093

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,* 1085–1114. http://dx.doi.org/10.1017/s030500091300028

Giezen, M. R., Escudero, P., & Baker, A. E. (2015). Rapid learning of minimally different words in five- to six-year-old children: Effects of acoustic salience and hearing impairment. *Journal of Child Language,* 1–28. http://dx.doi.org/10.1017/S0305000915000197

Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87,* B23–B34. http://dx.doi.org/10.1016/s0010-0277(02)00186-5

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B, 366,* 1179–1187. http://dx.doi.org/10.1098/rstb.2010.0321

Hawkins, S. (2010). Phonological features, auditory objects, and illusions. *Journal of Phonetics, 38,* 60–89. http://dx.doi.org/10.1016/j.wocn.2009.02.001

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28,* 377–396. http://dx.doi.org/10.1006/jpho.2000.0121

Henderson, A. M. E., Sabbagh, M. A., & Woodward, A. L. (2013). Preschoolers' selective learning is guided by the principle of relevance. *Cognition, 126,* 246–257. http://dx.doi.org/10.1016/j.cognition.2012.10.006

Hochmann, J. R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science, 14,* 1445–1458. http://dx.doi.org/10.1111/j.1467-7687.2011.01089.x

Højen, A., & Nazzi, T. (2015). Vowel bias in Danish word-learning: Processing biases are language-specific. *Developmental Science, 18,* 1–9. http://dx.doi.org/10.1111/desc.12286

Hollich, G. (2008). Supercoder: A program for coding preferential looking (Version 1.7.1) [Computer software]. West Lafayette, IN: Purdue University. Retrieved from http://hincapie.psych.purdue.edu/Splitscreen/

Jarvis, L. H., Merriman, W. E., Barnett, M., Hanba, J., & Van Haitsma, K. S. (2004). Input that contradicts young children's strategy for mapping novel words affects their phonological and semantic interpretation of other novel words. *Journal of Speech, Language, and Hearing Research, 47,* 392–406. http://dx.doi.org/10.1044/1092-4388(2004/032)

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology, 46,* 815–826. http://dx.doi.org/10.1037/a0019664

Krogh-Jespersen, S., & Echols, C. H. (2012). The influence of speaker reliability on first versus second label learning. *Child Development, 83,* 581–590. http://dx.doi.org/10.1111/j.1467-8624.2011.01713.x

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831–843. http://dx.doi.org/10.1038/nrn1533

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)* [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lmerTest

Labov, W., Ash, S., & Boberg, C. (2005). *Atlas of North American English: Phonetics, phonology, and sound change.* Berlin, Germany: Mouton.

Law, I. I. F., & Edwards, J. R. (2015). Effects of vocabulary size on online lexical processing by toddlers. *Language Learning and Development, 11,* 331–355. http://dx.doi.org/10.1080/15475441.2014.961066

MacRoy-Higgins, M., Shafer, V. L., Schwartz, R. G., & Marton, K. (2014). The influence of phonotactic probability on word recognition in toddlers. *Child Language Teaching and Therapy, 30,* 117–130. http://dx.doi.org/10.1177/0265659013487534

Mani, N., & Plunkett, K. (2007). Phonological specificity of consonants and vowels in early lexical representations. *Journal of Memory and Language, 57,* 252–272. http://dx.doi.org/10.1016/j.jml.2007.03.005

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11,* F9–F16. http://dx.doi.org/10.1111/j.1467-7687.2008.00671.x

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20,* 121–157. http://dx.doi.org/10.1016/0010-0285(88)90017-5

Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1376–1392. http://dx.doi.org/10.1037/0096-1523.22.6.1376

Medina, V., Hoonhorst, I., Bogliotti, C., & Serniclaes, W. (2010). Development of voicing perception in French: Comparing adults, adolescents, and children. *Journal of Phonetics, 38,* 493–503. http://dx.doi.org/10.1016/j.wocn.2010.06.002

Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development, 62,* 1288–1301. http://dx.doi.org/10.2307/1130807

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. *Child Development, 84,* 2064–2078. http://dx.doi.org/10.1111/cdev.12087

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition, 98,* 13–30.

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,* 522–537. http://dx.doi.org/10.1016/j.jecp.2008.05.003

Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical acquisition. *Cognitive Development, 22,* 271–279. http://dx.doi.org/10.1016/j.cogdev.2006.10.007

Quam, C., & Swingley, D. (2010). Phonological knowledge guides two-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language, 52,* 135–150. http://dx.doi.org/10.1016/j.jml.2009.09.003

Ramon-Casas, M., Swingley, D., Bosch, L., & Sebastián-Gallés, N. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology, 59,* 96–121. http://dx.doi.org/10.1016/j.cogpsych.2009.02.002

R Core Team. (2015). *R: A language and environment for statistical computing* (R version 3.2.0) [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Romeo, R. R., & Swingley, D. (2015, March). *Word recognition, phonological specificity, and SES: A longitudinal word-recognition study of toddlers*. Paper presented at the 2015 Biennial Meeting of the Society for Research in Child Development, Philadelphia, PA.

Schmale, R., Seidl, A., & Cristia, A. (2014). Mechanisms underlying accent accommodation in early word learning: Evidence for general expansion. *Developmental Science, 10,* 1–7. http://dx.doi.org/10.1111/desc.12244

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review, 120,* 779–797. http://dx.doi.org/10.1037/a0034191

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44,* 1321–1337. http://dx.doi.org/10.1044/1092-4388(2001/103)

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech, 46,* 265–294. http://dx.doi.org/10.1177/00238309030460021001

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science, 8,* 432–443. http://dx.doi.org/10.1111/j.1467-7687.2005.00432.x

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language, 60,* 252–269. http://dx.doi.org/10.1016/j.jml.2008.11.003

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76,* 147–166.

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

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology, 54,* 99–132. http://dx.doi.org/10.1016/j.cogpsych.2006.05.001

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language, 46,* 39–56. http://dx.doi.org/10.1006/jmla.2001.2799

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71,* 73–108. http://dx.doi.org/10.1016/s0010-0277(99)00021-9

Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language, 50,* 336–353. http://dx.doi.org/10.1016/j.jml.2003.11.004

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,* 1–30. http://dx.doi.org/10.1207/15250000252828226

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science, 21,* 221–226. http://dx.doi.org/10.1177/0963721412449459

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59,* 114–132. http://dx.doi.org/10.1016/j.jml.2008.03.001

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,* 362–378. http://dx.doi.org/10.1016/j.jml.2013.01.003

Yeung, H. H., & Nazzi, T. (2014). Object labeling influences infant phonetic learning and generalization. *Cognition, 132,* 151–163. http://dx.doi.org/10.1016/j.cognition.2014.04.001

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-old infants use distinct objects as cues to categorize speech information. *Cognition, 113,* 234–243. http://dx.doi.org/10.1016/j.cognition.2009.08.010

Yoshida, K., Fennell, C., Swingley, D., & Werker, J. F. (2009). 14-month-olds learn similar-sounding words. *Developmental Science, 12,* 412–418. http://dx.doi.org/10.1111/j.1467-7687.2008.00789.x

Appendix A

Effects of Experiment and CDI (z-Scored) on Likelihood of Manually Selecting the Familiar Object on Nonce Trials

Predictor	Coef.	Std.Err.	*z*-value	*p*-value
(Intercept)	−1.06	.35	−3.03	.002
Expt. 2	−.13	.38	−.35	ns
Expt. 3	−.35	.38	−.93	ns
z.CDI	−.35	.16	−2.18	<.03

Note. CDI = Communicative Development Inventory; Coef. = Coefficient; Std.Err. = Standard Error; Expt. = Experiment. Expt. 1 is used as the reference.

Appendix B

Fixed Effects in Analysis of Looking Responses Across Experiments

Predictor	Coef.	Std.Err.	*df*	*t*-value	*p*-value
(Intercept)	.733	.025	116.7	29.45	<.0001
Cond. MP	−.115	.026	71.0	−4.49	<.0001
z.CDI	.006	.015	116.7	<1	ns
Expt. 2	−.010	.036	116.7	<1	ns
Expt. 3	−.039	.035	116.7	−1.11	ns
Cond × CDI	−.036	.015	71.0	−2.31	<.03
Cond × Experiment 2	.039	.037	71.0	1.06	ns
Cond × Experiment 3	.029	.036	71.0	<1	ns

Note. Coef. = Coefficient; Std.Err. = Standard Error; Cond. = Condition; MP = mispronounced; CDI = Communicative Development Inventory; Expt. = Experiment. The dependent variable is each subject's per-condition mean proportion of looking to the target. Reference levels are Expt. = 1 and Cond. = CP.

(*Appendices continue*)

Appendix C

Fixed Effects in Item Analysis of Looking Responses Across Experiments

Predictor	Coef.	Std.Err.	df	t-value	p-value
(Intercept)	.720	.036	25.1	20.12	<.0001
Expt. 2	−.012	.044	31.6	−.26	ns
Expt. 3	−.048	.044	31.6	−1.11	ns
Cond. c-MP	−.087	.045	32.4	−1.95	= .06
Cond. v-MP	−.095	.046	33.6	−2.07	<.05
z.phonotactic	.040	.157	18.7	2.56	<.02
Expt. 2 × c-MP	.005	.062	31.6	.07	ns
Expt. 3 × c-MP	.004	.062	31.6	.06	ns
Expt. 2 × v-MP	.073	.062	31.6	1.19	ns
Expt. 3 × v-MP	.066	.062	31.6	1.07	ns

Note. Coef. = Coefficient; Std.Err. = Standard Error; Expt. = Experiment; Cond. = Condition; c-MP = consonant mispronunciation; v-MP = vowel mispronunciation. The dependent variable is the per-condition mean proportion of looking to the target, computed over subjects for the five target words. Reference levels are Expt. = 1 and Cond. = CP.

Appendix D

Fixed Effects in Analysis of Touching Responses Across Experiments

Predictor	Coef.	Std.Err.	z-value	p-value
(Intercept)	1.713	.314	5.45	<.0001
Expt. 2	.179	.397	.45	ns
Expt. 3	.560	.393	1.43	ns
Cond. MP	−.394	.262	−1.50	= .13
z.CDI	.214	.163	1.32	ns
z.phonotactic	.232	.111	2.08	= .04
Expt. 2 × MP	−.075	.371	−.20	ns
Expt. 3 × MP	−.938	.356	−2.63	<.01
z.CDI × MP	−.406	.152	−2.68	<.01

Note. Coef. = Coefficient; Std.Err. = Standard Error; Expt. = Experiment; Cond. = Condition; MP = mispronounced; CDI = Communicative Development Inventory. The dependent variable is the binary outcome of touching the familiar target object, or the other object, on a given trial. The model included random effects of subject and item. Reference levels are Experiment = 1 and Cond. = CP.