Listening to their peers: An assessment of toddlers' processing of other children's speech

by

Dana Bernier

A thesis

presented to the University Of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Linda Polka Professor School of Communication Science and Disorders McGill University
Supervisor:	Katherine White Associate Professor Department of Psychology University of Waterloo
Internal Members:	Stephanie Denison Associate Professor Department of Psychology University of Waterloo
	Elizabeth Nilsen Associate Professor Department of Psychology University of Waterloo
Internal-external Member:	Svetlana Kaminskaïa Associate Professor French Studies University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Bernier, D. E., & White, K. S. (2019a). Toddlers' sensitivity to phonetic detail in child speech. *Journal of Experimental Child Psychology*, *185*, 128–147. https://doi.org/10.1016/j.jecp.2019.04.021

D. E. B. created the initial design, programmed the experiments, collected data, performed data analyses, developed the initial interpretations of the results, and wrote and revised the initial draft of the manuscript. K. S. W. provided critical feedback on the design of the study and interpretation of results, as well as revisions to both the original manuscript and later versions for publication.

Bernier, D. E., & White, K. S. (2019b). Toddlers process common and infrequent childhood mispronunciations differently for child and adult speakers. *Journal of Speech, Language, and Hearing Research*, *62*, 4137-4149. https://doi.org/10.1044/2019_JSLHR-H-18-0465

D. E. B. created the initial design, programmed the experiments, collected data, performed data analyses, developed the initial interpretations of results, and wrote and revised the initial draft of the manuscript. K. S. W. provided critical feedback on the design of the study and interpretation of results, as well as revisions to both the original manuscript and later versions for publication.

Abstract

Young language learners acquire their first language(s) from the speech they are exposed to in their environment. For at least some children (e.g., those in daycare), this environmental speech includes a large quantity of speech from other children. Yet, we know little about how young learners process this type of speech and its status as a source of input. This dissertation will examine these questions by assessing toddlers' processing of older children's speech, both in general and as a function of their experience with other children. Chapter 2 begins by showing that 21- to 23-month-olds processed a 7-year-old speaker's productions of familiar words as well as those of an adult, and with the same level of sensitivity to phonetic detail previously shown for adult speakers. This is followed in Chapter 3 with 21- to 23-month-olds showing differential treatment of mispronunciations of familiar words depending on the type of mispronunciation (common to child speech or rare), the speaker that produced them (an adult or a 6-year-old), and the specific contrast involved (/w/-/ɹ/ or /t/-/ʧ/). Finally, Chapter 4 shows that by 3 years at least some children may have the expectation that a 5-year-old would mispronounce /ɹ/ as [w] (e.g., rainbow as wainbow), but that an adult would not. Effects of experience were assessed across all experiments. When effects of experience with other children were found, they were small. Overall, these findings demonstrate that child speech may represent useful input for young language learners.

Acknowledgements

I would like to thank Dr. Katherine White for her support and expertise in the completion of these studies, for her advice, and for being a friend. Thanks to the other supervisory committee members (Dr. Stephanie Denison, Dr. Liz Nilsen, Dr. Linda Polka, and Dr. Svetlana Kaminskaïa) for their helpful suggestions and insights in the development of this final document. I would also like to thank the LIDL lab members who helped in some way or another with the completion of these studies, including Emily McIntosh, Joel LeForestier, Shaquille Sealy, Erin Kim, Ashley Blayney- Hoffer, Ayah Taji, Catherine Rocha, Chloe Whitham, Shirley He, Kristen Thompson, and Annika Voss. I would also like to thank the many families who volunteered their time to participate. Finally, I thank my family: my father (André) and my mother (Lynne), my brother (Mike), sister-in-law (Leslie) and their children (Delia, Sophie, and Charlie), my sister (Giselle), brother-in-law (Cory) and their children (Emma, Ayla, and Damien), my youngest brother (Ryan), my grandparents and many aunts and uncles, and most significantly my partner Doug, for their endless support and assistance, and for their behind the scenes contributions to my being able to get to this point. This research was supported by a Post Graduate Scholarship from the Natural Sciences and Engineering Council of Canada.

Table of Contents

List of Figures

List of Tables

Chapter One: General Introduction

Language is something that children the world over acquire, making this amazing feat seem a simple task. The human predilection towards communication begins early and helps drive children's acquisition of language. The fact that there are roughly 6,500 different languages spoken today is a testament to the diversity of this very human form of communication. Which of these many languages a child will learn in their first few years of life is determined by which language(s) they are exposed to – both at home and in the community at large.

Moving beyond which language a child will learn, each child's experience with that language is also important in determining their language abilities (Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015). Differences in the amount of speech a child hears in a specific language (e.g., a 3-year-old with affluent parents is likely to have heard 30 million more words directed at them – and in richer contexts - than a same-aged peer from a less affluent home; Hart & Risley, 1995) as well as differences in how representative that speech is of the larger community (e.g., whether the speech is from a native or a non-native speaker of that language) will impact how children understand and use that language. Thus, both the quantity and quality of what children hear (their speech input) will influence their language knowledge and processing abilities.

To date, the majority of the work examining the influence of children's speech input on their language outcomes has focused on the influence of native adult speakers – in particular female adults, who are considered children's primary source of language input. However, young children's language environments include a variety of other speakers as well. And for many young language learners, these speakers include other children (e.g., siblings, peers in daycare). Although other children represent a large part of the language experience of many young children, we know very little about how their speech is processed, and its status as a source of language input.

With this dissertation, I hope to help fill this gap. In this introduction, I will first briefly review some of what we know about the potential problems posed by children's speech, and what little we know of how young children process it, contrasting this with what we know about their processing of adult speech. I will then discuss the potential role that experience with other children may have on young language learners' processing of child speech, using examples of how experience has been proposed to affect other types of speech processing. Finally, I will introduce three sets of experiments examining toddlers' processing of words produced by children.

The Productive Capabilities of Young Children

Prior to infants' first words, they have been vocalizing for months, turning into babbling (CV (consonant-vowel) sequences) around 6 months. This early vocalization and babbling occurs independently of which language a child is learning (Ingram, 2008). These early productions are also unaffected by such factors as socio-economic status (SES) and hearing impairment (Hoff, 2006), with all infants (even the congenitally deaf) producing speech-like sounds.

However, by about 8-10 months of age, infants' babbling shows increases in complexity, moving from reduplicated syllables to more complex and variable combinations of syllables, and it begins to take on a conversational tone, with the stress and intonation patterns of the ambient language (Stoel-Gammon & Vogel Sosa, 2007). Infants also begin to include some new language-specific syllable shapes, such as CVC for English-learning children (Stoel-Gammon & Vogel Sosa, 2007).

Around 12 months, infants' first words start to appear, though they do not yet take on an adult-like form. These first words co-occur with infants' later babbling and use the same consonants (primarily stops, nasals, and glides) as well as the same set of simple syllabic structures (CV as in *go*, CVC as in *sit*, and CVCV as in *baby*; Stoel-Gammon & Vogel Sosa, 2007). By around 18 months, infants have a productive vocabulary size of about 50 words. One notable feature of early word productions is how different they sound from the adult target words.

In the early word productions of typically developing children, sounds that are infrequent in the language and/or those that are difficult to articulate are often missing (Stoel-Gammon & Vogel Sosa, 2007). The order in which these sounds are first produced (emergence), and the age at which children reach an adult-like production (accuracy), appear to be dependent on three factors – articulatory complexity, functional load, and ambient frequency (Ingram, 2008). The first factor, articulatory complexity, refers to how much precision is required to be able to produce a given sound (i.e., how difficult it is to produce). The second factor, functional load, represents the relative importance of a sound within the language's phonological system. Put simply, it refers to the number of different words a particular sound occurs in. This is distinct from the third factor, ambient frequency, which represents how often a sound is heard. For example, /ð/ is a sound that occurs in very few English words (e.g., the, this, that). It would therefore have a low functional load. However, due to its occurrence in function words which are produced quite frequently, /ð/ would have a high ambient frequency. For English-learning infants, emergence is strongly influenced by functional load, whereas accuracy is strongly influenced by articulatory complexity (Stokes & Surendran, 2005).

Starting at around 18 months, there is a rapid increase in vocabulary size, reaching 200-300 words by the time children are 2 years old, and a few thousand words by the time they are school age. Along with this increase in the number of words children are producing, they are also expanding their repertoire of speech sounds. The speech of 2-year-olds already contains the basic elements of the adult language; for English-learning children this includes labial and alveolar stops and nasals (/p b m t d n/), as well as glides (/w j/). But it is still restricted to simple CV or CVC syllables and their combinations (Stoel-Gammon & Vogel Sosa, 2007). Though this repertoire works well for some words, there are still many words that young children know that fall outside of their productive capabilities. For these words, children will systematically alter their productions to fit within their own productive capabilities. These phonological processes include (among others) substitutions of one sound for another (e.g., *fumb* for *thumb*) and simplification processes that involve omissions of sounds (or syllables) altogether (e.g., *nake* for *snake* or *nana* for *banana*). These systematic deviations are likely to involve replacing sounds with a low functional load with sounds that have a high functional load (e.g., /f/ has a higher functional load than /θ/), and also tend to be produced earlier (Edwards, Beckman, & Munson, 2015). By the time children are 4 years old, their phonetic inventories will have grown to include nearly all of the consonants in the English language, as well as a much broader range of syllable types (e.g., CCV as in *tree*, CCVC as in *grass*, and CCCV as in *straw*; Dodd, Holm, Hua, & Crosbie, 2003; Stoel-Gammon & Vogel Sosa, 2007).

Even though children's phonetic inventories are nearly complete at this point, the accuracy of their productions progresses much more slowly. By the time children are 4, the majority of the larger deviations described above have decreased (Dodd et al., 2003). But even 6-year-olds, whose speech contains very few of the larger deviations seen in younger children, continue to show less accuracy and more variability in their productions than adults. For example, children this age produce less accurate /l/ (Lin, Inkelas, McDonnaughey,

& Dohn, 2016), more variable /s/ (Koenig, Lucero, & Perlman, 2008; Munson, 2004), less distinguishable /s/ and /ʃ/ (Maas & Mailend, 2017; Nissen & Fox, 2005), and have a larger vowel space than adults (Hillenbrand, Getty, Clark, & Wheeler, 1995; Lee, Potamianos, & Narayanan, 1999). These deviations and increased variability mean that even older children's speech may be more difficult for a young language learner to process than adult speech.

The Word Processing Abilities and Lexical Representations of Young Children

Before addressing young children's processing of other children's speech, it is necessary to establish a baseline against which a comparison can be made. As such, I will begin by discussing young children's processing of adult speech.

Speech from Adults

Along with young children's vocabulary burst across the latter half of their 2nd year come rapid gains in the efficiency with which they process familiar words (Fernald, Perfors, & Marchman, 2006; Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998; Hurtado, Marchman, & Fernald, 2007). These gains have been demonstrated in looking tasks, which show that at 15 months (a few months prior to the 50-word milestone) toddlers need to hear a full target word before orienting towards the correct object. But by the time children reach 24 months, they are already shifting their gaze to the correct object well before the speaker has finished producing the target word. In other words, they are already showing an important hallmark of mature speech processing – that of incremental processing of the input.

The mental representations of these familiar words have been proposed by some to initially not be well specified in very young children (e.g., Charles-Luce & Luce, 1990). However, more recent work has shown that, during word processing, toddlers are in fact very sensitive to the kinds of phonetic changes that signal differences in meaning (e.g., that is, to changes like /b/ to /p/, which results in a change of the word *bat* to the word *pat*). This suggests that toddlers not only encode words with this level of phonetic detail, but also that they use this detail during word processing.

To assess the level of detail in toddlers' lexical representations, researches have looked at toddlers' sensitivity to mispronunciations of familiar words (e.g., *vaby* for *baby*). The premise is that if toddlers' lexical representations contain detailed information about the characteristics of the speech sounds that make up the word, they should show a significant reduction in looking or accuracy for a mispronounced label, given that it does not match their lexical representation. However, if their lexical representations are not well specified (as suggested by Charles-Luce & Luce (1990) and others), then toddlers should treat the mispronounced label as an acceptable label.

This work has shown that toddlers detect even slight mispronunciations of familiar words produced by adults (e.g., Swingley & Aslin, 2000). For example, in looking paradigms toddlers will spend less time looking towards an image of a baby when it is labelled as *vaby* compared to when it is labelled correctly. In addition to general processing costs for mispronunciations, toddlers also show graded sensitivity to the degree of mispronunciation (Mani & Plunkett, 2011; Ren & Morgan, 2011; Tamási, McKean, Gafos, Fritzsche, & Höhle, 2017; White & Morgan, 2008). Toddlers look progressively less to the image of a familiar object as the degree of mispronunciation increases (e.g., White & Morgan, 2008). For example, there may be a relatively small cost associated with a small phonetic change, but a progressively larger cost associated with progressively larger changes. This graded sensitivity has also been demonstrated using pupillometry, where the larger the degree of mispronunciation, the greater the change in pupil diameter (Tamási et al., 2017).

However, in some studies, asymmetries have been found for infants' and toddlers' processing of different types of mispronunciations. For example, some studies (Altvater-Mackensen & Fikkert, 2010; Altvater-Mackensen, van der Feest, & Fikkert, 2014; Nam & Polka, 2016; van der Feest & Fikkert, 2015) report that infants and toddlers detect single-feature phonetic changes in one direction (e.g., /b/→/d/, /v/→/b/), but not in the opposite direction (e.g., /d/→/b/, /b/→/v/). These asymmetries are often (though not always) consistent with differences in input frequency or with underspecification accounts of lexical representations. Input frequency account predictions are based on a sound's functional load and/or its ambient frequency in the input. Underspecification account predictions, on the other hand, are based on some sounds being the default or unspecified sound (e.g., the coronal stop /d/), with other similar sounds being marked to indicate how they differ from the default/unmarked sound (e.g., /b/ is marked as labial relative to the coronal /d/; /v/ is marked as fricative relative to the stop /b/). Because frequency and markedness status are often correlated, both of these accounts predict that listeners are more likely to detect changes to more frequent/unmarked sounds (e.g., coronals, stops) than the reverse (e.g., labials, fricatives).

Speech from Other Children

There are very few existing studies on young children's processing of child speech. In an early study, Dodd (1975) tested 2- to 4-year-olds' comprehension of their own productions of object labels, as well as those of another child, and an adult. The object labels produced by each child were those that the child consistently mispronounced, and thus contained large phonological deviations (e.g., substitution or omission of consonants, omission of unstressed or initial syllables, or the addition of sounds). In some cases, more than one of these deviations occurred in the same word. For example, *beya* [bejʌ] for *umbrella* contains 3 errors – the initial unstressed syllable /ʌm/ is omitted, the /ɹ/ sound is omitted from the second syllable, and the /l/ in the third syllable is substituted with a [j]. Overall, Dodd found that children were less accurate in choosing the named object when hearing labels produced by a child (whether their own productions or that of another child) compared to labels produced by an unfamiliar adult. Further examination of children's accuracy for the child productions showed that the words they had the most difficulty with contained significantly more phonological deviations per syllable than the words they had the least difficulty with.

More recently, Cooper, Fecher, and Johnson (2018) used a looking paradigm to test the same question with 2½-year-olds, and similarly found an advantage for adult speech. In this study, toddlers were presented with the images of two familiar objects on a screen, along with auditory labels of one of the familiar objects. The labels children heard during the test session were previously recorded, and produced by either familiar speakers (the child, the child's mother) or unfamiliar speakers (a same-gender peer, the mother of the peer). Given the age of the children in this study, it is likely that the children's productions contained many phonological deviations, much like those in Dodd (1975). Cooper et al.'s (2018) results show that children's looking behaviour did not differ between familiar and unfamiliar speakers, and overall children looked less to the labelled object for the child speakers than adult speakers. Together with the results of Dodd (1975), these studies suggest that young children have difficulty processing speech that deviates significantly from adult target forms, regardless of who produced it. Cooper et al. (2018) additionally used their results to argue that children's lexical representations take the forms produced by adults, rather than the forms of the children's own deviant productions.

In a recent study, Krueger, Storkel, and Minai (2018) tested 4- and 5-year-old children's perception of correct and naturally occurring mispronounced labels produced by other children using a variety of different types of errors (e.g., thumb pronounced correctly or as *fumb*). Using a computerized selection task, Krueger et al. (2018) found that the preschoolers were accurate in choosing the named object for both correct and mispronounced labels (93% and 87% of the time, respectively). However, they also showed a processing delay for the mispronounced labels, being slower to select the named object with less direct mouse movements; this finding was mirrored with adult speech in a separate experiment.

In other work examining the precision with which slightly older children's speech is processed, Strömbergsson, Wengelin, and House (2014) examined 4- to 6-year-old's processing of their own and other children's synthetically modified speech. Children (typically developing and those with a phonological disorder) were presented with the image of a familiar object (Experiment 1) or a child (Experiment 2) and the image of an 'X' on a screen, along with audio labeling an object. These labels were modified by replacing the onset with either a different version of the same sound (i.e., one [t] replaced with another [t]), or a different sound altogether (i.e., a [t] replaced with a [k]). Children were asked to evaluate the accuracy of the production and choose the object/child if they perceived the word to be correct and the 'X' if they perceived it to be wrong. Strömbergsson et al. found that typically developing children had little difficulty distinguishing between the correct and mispronounced versions, regardless of whether the speech was their own or that of another child. And while children with a phonological disorder were found to have some difficulty processing their own speech (at least without a time delay), they did not show this difficulty when processing the speech of other children. Thus, by 4 to 6 years, children are sensitive to even small pronunciation changes in the speech of similarly aged children.

These studies provide some indication of how young children process the speech of their age-matched peers. In particular, they show that toddlers have difficulty with the speech of other toddlers, while pre-schoolers fare much better with the speech of other pre-schoolers.

However, age-matched peers are not the only type of child-child interaction possible. Given that child speech constitutes a large percentage of the input for some young language learners, it is important to continue exploring how toddlers process the speech of children of various ages, the factors that might affect ease of processing, and the extent to which speech from other children is used to guide language learning.

The Role of Experience in Processing Atypical Speech

As the above review describes, the quality of children's speech is not equal to that of adults', yet by the time children reach pre-school age, they are able to cope with small deviations in their peers' speech and thus have little difficulty understanding them. Because pre-schoolers are older than toddlers, they are also likely to have had a greater amount of experience with child speech. Given that experience has been shown to help adult listeners process child speech more effectively (Munson, Johnson, & Edwards, 2012), my dissertation explores the possibility that toddlers' ability to process child speech is based on how much experience they have had hearing other children speak. A parallel may be found in work on the effects of experience on children's processing of other types of non-canonical speech. Just like child speech, accented speech – whether it's a different dialect or a non-native accent – is speech that differs from the standard form.

This work has shown that before 20 to 24 months, toddlers have difficulty processing unfamiliarly accented speech (Best, Tyler, Gooding, Orlando, & Quann, 2009; Mulak, Best, Tyler, Kitamura, & Irwin, 2013; Newman, Morini, Kozlovsky, & Panza, 2018; Schmale, Cristia, & Seidl, 2012; Schmale, Seidl, & Cristia, 2015; van Heugten & Johnson, 2014; van Heugten, Krieger, & Johnson, 2015; White & Aslin, 2011). For example, Best et al. (2009) with American toddlers and Mulak et al. (2013) with Australian toddlers, used different tasks, yet both found that 15-month-olds were only able to recognize familiar words produced in their native English accent, and not words produced with an unfamiliar Jamaican-English accent.

Importantly, some of this work has shown that exposure to the specific unfamiliar accent improves toddlers' processing of speech in that accent. For example, White and Aslin (2011) found that providing 18- to 20-month-olds with examples of the way an accented speaker produced a familiar label (e.g., that she pronounces dog as *dag*), allowed toddlers to recognize the accented speaker's intended referent for similarly accented words (e.g., bottle pronounced as *battle*). Likewise, van Heugten and Johnson (2014) found that 15-month-olds were able to recognize familiar words that were produced with an unfamiliar accent, after they had been exposed to the accented speaker reciting a familiar story. Finally, Schmale et al. (2012) found that exposure to passages spoken in a particular foreign accent improved 24-month-olds' processing of novel words from the same accented speaker.

Exposure to an accent has been suggested to help both adults and children in a couple of ways: (1) by allowing them to learn the specific pronunciation changes of the particular accent, and (2) by increasing their tolerance for atypical pronunciations.

Learning Specific Speech Patterns

One way in which listeners appear to cope with challenging or unusual speech is by learning the specific deviations the speaker makes relative to the listener's native/familiar accent. A large body of work has shown that listeners of all ages are capable of learning specific patterns of deviation via lexically guided learning (adults: Kraljic & Samuel, 2011; Kraljic, Samuel, & Brennan, 2008; Norris, McQueen, & Cutler, 2003; toddlers and children: McQueen, Tyler, & Cutler, 2012; Newman et al., 2018; White & Aslin, 2011).

In adults, lexically guided adaptation is typically demonstrated in a paradigm in which listeners are exposed to ambiguous consonant sounds embedded in words. Following this exposure, listeners' judgements of the category boundaries for speech sounds adjust to accommodate the atypical sounds. For example, adults who are exposed to the /s/ (or /ʃ/) in words replaced with an ambiguous [sʃ] sound (i.e., giving the speaker an atypical production pattern), would later include the ambiguous [sʃ] sound within the /s/ (or /ʃ/) category for that speaker (Kraljic et al., 2008). Importantly, this type of accommodation has been shown to occur only when this variation is characteristic of the speaker, and not when it occurs incidentally (e.g., the speaker has a pen in their mouth). Similar results have been demonstrated for children as young as 6 years (McQueen et al., 2012). This demonstrates that both adults and children are able to learn and adapt to specific deviations in speakers' productions.

This kind of adaptation is also seen in even younger listeners (White & Aslin, 2011). As described above, after being exposed to an unfamiliar accent in which *dog* was pronounced as *dag*, toddlers not only learned the way the speaker pronounced the words they heard, but also generalized the pronunciation to words they never heard the speaker pronounce (e.g., that is, they inferred that *battle* referred to *bottle*). Importantly, this adaptation was shown to be specific to the accent the child heard, and not to any modification (e.g., neither *deg* nor *bettle* were treated as valid labels for dog and bottle).

Once learned, listeners can invoke knowledge of these specific patterns in response to cues. For example adults who were primed with a speaker's nationality altered their speech categorization to match the nationality they were primed with (Hay & Drager, 2010; Hay, Nolan, & Drager, 2006; Niedzielski, 1999). Thus, not only are these kinds of adaptations effective in the short-term, but they can also be stored and used later when a similar situation is encountered.

Increased Tolerance for Atypical Speech

An alternative way people may cope with atypical speech is by increasing their tolerance for it. This strategy would require listeners to be more accepting of errors or deviations from the expected patterns. This openness to accommodating atypical speech could occur due to adaptations based on low-level perceptual information, or based on the use of higher-level knowledge.

Greater Variability of the Speech Signal. Perceptual information in the form of more diverse acoustic input has been proposed to shift listeners' tolerance towards accepting atypical forms of speech. In toddlers, this processing strategy was demonstrated in 24-month-olds exposed to highly variable speech in the form of a foreign accent (Schmale et al., 2012). In this study, children were trained to associate a novel label with a novel object by a native English speaker, and tested on their recognition of this object-label mapping by a Spanish-accented speaker. Prior to this, half of the children were familiarized to the native English speaker, and the other half to the Spanish-accented speaker. Children who were first familiarized to the Spanish-accented speaker appeared to show more robust learning than those familiarized to the native English speaker. Thus, hearing an acoustically variable foreign accent helped children process test items that were produced in the same foreign accent.

Similar results were also found when 24-month-olds were familiarized to four different speakers who varied in age and sex, but who had the same native accent as the children (Schmale et al., 2015). Following the exposure to this acoustically variable native speech, the 24-month-olds were able to recognize newly learned words produced with a non-native Spanish accent. Thus, as with exposure to the foreign accent above, exposure to more acoustic variability in their native accent appeared to help children process test items produced in a foreign accent.

General Expectations. Alternatively, a listener's tolerance for atypical speech could increase because of the listener's prior knowledge about a speaker's likelihood of making mistakes (e.g., a foreign sounding individual may be more likely to make grammatical errors). This strategy has been shown to be used by both adults (Gibson, Bergen, & Piantadosi, 2013; Gibson et al., 2017; Lev-Ari, 2015; McQueen & Huettig, 2012) and children (Rett & White, 2019; Yurovsky, Case, & Frank, 2017).

This type of processing strategy has been explored primarily for the processing of non-native speech, and has been shown to occur when the speech contains noise, such as grammatical errors. Indeed, adults appear to expect grammatical errors from non-native speakers (Hanulíková, van Alphen, Goch, & Weber, 2012), and so correct for them. For example, adults were more likely to correct the word order of implausible utterances such as "The mother gave the candle the daughter" to arrive at plausible interpretations when they were spoken by non-native speakers (Gibson et al., 2017). Additionally, when adults were exposed (via filler material) to a high proportion of sentences containing syntactic errors (such as a function word being deleted or inserted, as is common of non-native speakers), they were more likely to correct implausible test sentences (Gibson et al., 2013) suggesting that the listeners' strategy was primarily determined by their prior real-world experience with native and non-native speakers. In addition to correcting for grammatical errors, adults have also been shown to rely more heavily on context to determine the intended referent for a non-native speaker (Lev-Ari, 2015). Together, these findings show that when faced with speakers who are expected to make errors, adults will both correct for these errors and rely more on context to determine a speaker's intended referent.

Recent work with young children suggests that they too are able to flexibly use expectations about a speaker during language processing (Rett & White, 2019; Yurovsky et al., 2017). For example, 4- to 6-year-olds are more likely to ignore grammatical errors (e.g., "That am a shoe") from a non-native speaker than a native speaker in a speaker reliability task (Rett & White, 2019). And, just like adult listeners, child listeners' prior experience with a speaker (whether they produced plausible or implausible utterances) determined how likely they were to infer a plausible meaning for an implausible utterance (e.g., *bees* in the utterance "I had carrots and bees for dinner" was more likely to be interpreted as *peas* when it came from a speaker who had previously produced plausible utterances; Yurovsky et al., 2017). Even though 4- and 5-year-olds were shown to use their expectations in the same way adults do, they were also more likely than adults to make a plausibility-based inference regardless of speaker, demonstrating a greater reliance on world knowledge (e.g., that you don't eat bees for dinner). Importantly, 3-year-olds did not show differences across speakers, suggesting that they were either not developing speaker-specific expectations or not using them to guide their processing. However, it is possible that a more sensitive measure, such as looking time, would reveal this ability.

What This Suggests for the Processing of Older Children's Speech

As the above review shows, exposure to accents can help listeners (including toddlers) process accented speech. In the same way, it is possible that hearing other children speak about familiar things may subsequently help toddlers process the speech of an unfamiliar child. The above review also shows that there are multiple routes through which experience can facilitate the processing of atypical speech. Applying this to child speech, first, exposure to the way children produce speech sounds, including the types of errors children typically make, could help toddlers adapt to the speech of young children. For example, experience hearing speech sound replacements could allow them to adapt to the deviant sounds by linking them to the appropriate category (e.g., *bwush* interpreted as *brush* because [w] is treated as /ɹ/, by virtue of it being a common substitution in young children's speech).

Second, experience listening to other children speak could lead toddlers to pay less attention to the specifics of child speakers' productions (e.g., *bwush* could be interpreted as *brush* because the word is produced by a child and it sounds something like the label of the object the child is interacting with). Toddlers may increase their tolerance in this way either because of the high variability of the child speaker's productions or because they rely on their general knowledge about child speakers (that they will make errors).

The Kinds of Child Experience that Might Matter

The above review suggests that experience may influence toddlers' processing of children's speech. If so, environments where there is a high amount of child-to-child speech will be those most likely to provide toddlers with this experience. Although older siblings might be considered a major source of experience with child speech, siblings are not the only children young language learners are exposed to in their daily lives. Many children also encounter other children in daycares, as well as in formal and informal community activities such as swimming, gymnastics, community drop-in centres, or even neighbours and family friends. With dual-income families that have 1 to 2 children being common (Bernard, 2018), these other sources of child-child experience are likely to be very important in the lives of many young language learners.

When comparing daycares to home environments, there are clear differences in both the amount of time children engage with other children, as well as the number of children they engage with. While home environments for many people in Western cultures are likely to have only one adult caring for one or two children, daycare environments typically have more adults, each caring for two to five children. This difference in the overall number of children could affect toddlers in a couple of ways. The first is the amount of time toddlers spend interacting with other children. In fact, when compared to children cared for at home with a sibling, children in daycares have been shown to spend more time engaged in verbal interactions with other children (Soderstrom, Grauer, Dufault, & McDivitt, 2018). The second is the diversity of child speakers that toddlers engage with. Exposure to more variable input has been shown to increase infants' attention to phonetic detail during word learning (Rost & McMurray, 2009, 2010) as well as increase their learning of a novel phonotactic pattern (Seidl, Onishi, & Cristia, 2014), and therefore could also influence toddlers' learning of child speech patterns. Together this suggests that toddlers who regularly experience group settings (whether in daycare or some other activity) may be those most likely to develop knowledge of the way other children speak, regardless of whether or not they have siblings. There are, of course, other co-occurring features in many (but not all) group settings that may also impact children's language knowledge or processing (e.g., the greater variety of adult speakers, or the number of structured interactions). Despite this, the amount of time toddlers spent with other children in daycare, informal playgroups, and organized activities was used as a measure of experience in this dissertation.

Current Studies

In the current series of studies, young children's perceptions of the speech of other children will be examined. Given that both toddlers (Cooper et al., 2018; Dodd, 1975) and naïve adult listeners (Flipsen, 2006; Hodson & Paden, 1981) find the speech of very young children challenging to understand, this dissertation will explore toddlers' processing of slightly older children's speech. In particular, I focus on the speech of early school-aged children (5-7-year-olds). This type of speech, though less dramatically different from canonical adult speech, is still noisier and less accurate than adult speech. In keeping with previous work, only female speakers were used.

The first goal of the present dissertation is to determine how well toddlers process the speech of a slightly older child. I therefore begin in the next chapter by establishing toddlers' basic processing of familiar words from an early school aged child. This is followed by a more fine-grained analysis of their perception of these familiar words. In both experiments of Chapter 2, the role of toddlers' experience with other children will be considered. One possible outcome is that toddlers will have difficulty processing the child speech. In particular, they may show less fine-grained processing of child speech compared to adult speech. If this is the case, toddlers with more experience hearing children speak may show better processing than those with less experience.

A second goal of the present dissertation is to determine whether toddlers have expectations about what other children sound like – either generally as speakers who mispronounce words, or more specifically as speakers who make certain kinds of errors. Chapter 3 begins this inquiry by examining whether toddlers process familiar words differently depending on whether they are mispronounced in ways that are common in child speech versus ways that are rare, and whether the speaker – a child or an adult – matters. If toddlers have specific expectations about the types of errors made by child speakers, then they should accept commonly produced labels as valid labels for familiar objects when they are produced by a child speaker, but not when they are produced by an adult. In contrast, no such difference should be seen for rare mispronunciations, and they should be rejected as labels from both speakers. And again, toddlers with more experience with other children may be more likely to show this pattern.

In Chapter 4, toddlers' expectations about how children speak will be examined in a word learning task. Toddlers will be trained on a new word-object mapping by an adult and subsequently hear a mispronunciation of this word from a child speaker. Unlike highly familiar words, toddlers will have never previously been exposed to mispronunciations of this word. However, if toddlers have expectations about the way children speak, then they should infer that the intended referent of the child's mispronounced label is the trained object. In contrast, because native-speaking adults do not typically make these same kinds of errors, toddlers should infer that an adult using the mispronounced label is producing a new word that should be mapped to a novel (never-before-seen) object. And, as with the other chapters, if toddlers do have these expectations, then those with more experience hearing the speech of other children should be better equipped to understand the intended referent of a child speaker.

Chapter Two: Sensitivity to Phonetic Detail in Child Speech

The following chapter has been published as:

Bernier, D. E., & White, K. S. (2019a). Toddlers' sensitivity to phonetic detail in child speech. *Journal of Experimental Child Psychology*, *185*, 128–147. https://doi.org/10.1016/j.jecp.2019.04.021

Young language learners acquire their first language(s) through exposure to speech in their environment. Although this environment is typically thought of exclusively in terms of adult speech, language learners are also exposed to the speech of other children. In some environments, this child speech may even occur in an amount similar to that of adult speech (Bernier & Soderstrom, 2016). Yet we know very little about how well young children process this kind of speech. Given that other children's speech potentially represents a substantial source of input for at least some children (e.g., those with siblings or attending daycare), this represents a significant gap in our understanding of early language processing.

Very young children's productions are characterized by a number of phonological deviations from adult targets. These changes include (among others) substitutions of one sound for another (e.g., *fumb* for *thumb*) and omissions of sounds (or syllables) altogether (e.g., *nake* for *snake*) and can lead to low intelligibility for naïve adult listeners (Flipsen, 2006; Hodson & Paden, 1981). The majority of these more significant phonological deviations have decreased by 4 years of age, and they are largely absent after 6 years (Dodd et al., 2003). These later productions are (on average) quite intelligible for even naïve adult listeners, although adults with more experience listening to children are more accurate and reliable in their judgments of individual sounds (Munson et al., 2012). However, even once these larger deviations are no longer present, children continue to show less accuracy and more variability in their productions than adults. For example, early school-aged children produce less accurate /l/ (Lin et al., 2016), more variable /s/ (Koenig et al., 2008; Munson, 2004), and less distinguishable /s/ and /ʃ/ (Maas & Mailend, 2017; Nissen & Fox, 2005; R. Romeo, Hazan, & Pettinato, 2013) than adults, and they have a larger vowel space (Hillenbrand et al., 1995; Lee et al., 1999). These deviations and increased variability mean that child speech may be more difficult for young language learners to process than adult speech.

In an early study exploring young children's processing of child speech, Dodd (1975) tested 2- to 4- year-olds' comprehension of their own productions of object labels as well as those of another child and an adult. She found that children were less accurate in choosing the named object when hearing labels produced by a child (including their own productions) compared with labels produced by an adult. More recently, Cooper, Fecher, and Johnson (2018) used a looking paradigm to test the same question with 2½-year-olds and similarly found an advantage for adult speech. These studies suggest that young children have difficulty processing speech that deviates significantly from adult target forms. In other work more closely examining the precision with which child speech is processed, Strömbergsson, Wengelin, and House (2014) examined slightly older (4- to 6-year-old) children's processing of their own and other children's synthetically modified speech. They found that typically developing children were able to distinguish between correct and mispronounced versions of resynthesized speech that involved changes between /t/ and /k/ regardless of whether the speech was their own or that of another child. Although children with a phonological disorder were found to have some difficulty processing their own speech (at least without a time delay), they did not show this difficulty when processing the speech of other children. Thus, by 4–6 years of age, children are sensitive to even small pronunciation changes in the speech of similarly aged children.

These studies provide some indication of how young children process the speech of their age-matched peers. Namely, on average, toddlers have difficulty processing the speech of other toddlers, but preschool-aged children seem to fare much better in processing the speech of other preschool-aged children. However, age-matched peers are not the only type of child–child interaction possible. Given that child speech constitutes a large percentage of the input for some young language learners, it is important to continue exploring how toddlers process the speech of children of various ages, the factors that might affect ease of processing, and the extent to which speech from other children is used to guide language learning.

Although there is limited work to date on toddlers' processing of child speech, there is much more work on their processing of other forms of noncanonical speech, such as non-native accented speech. Although there are processing costs associated with initial exposure to an unfamiliar accent (e.g., Best, Tyler, Gooding, Orlando, & Quann, 2009; van Heugten & Johnson, 2014; White & Aslin, 2011), continued exposure to the accent can lead to successful word recognition (Schmale et al., 2012; van Heugten & Johnson, 2014; van Heugten et al., 2015; White & Aslin, 2011). These studies suggest that if child speech is initially difficult for toddlers to process, experience hearing other children speak may help.

It is clear that toddlers hear their own productions (and so have experience with child speech). However, the role of experience with *other* children has not previously been considered. Experience hearing other children's speech could facilitate processing in a number of ways. One way that experience may help is by allowing listeners to learn something about the sound categories children produce. Just as exposure to ambiguous sounds embedded in words alters adults' judgments of phoneme category boundaries (Norris et al., 2003), exposure to the speech of other children could refine toddlers' perception of those speech productions. In this case, it is expected that toddlers with greater exposure to other children would show *more* sensitivity to phonetic detail in child speech.

Alternatively, experience with children could lead toddlers to pay less attention to the specifics of child speakers' productions and rely more on context to determine referential intent. It is clear that adult listeners rely heavily on context for comprehending very young children (whose extreme phonological deviations make them highly unintelligible). And recent work has shown that even young children will use top-down information during word processing when the acoustic input is unreliable (Yurovsky et al., 2017). If toddlers have knowledge that children misarticulate sounds, it is possible that they would adopt this strategy in the case of child speech. In this case, it is expected that toddlers with greater exposure to other children would be more tolerant of deviations (and, as a result, show *less* sensitivity to phonetic detail) in child speech.

Regardless of how experience might influence processing of children's speech (either by affecting perception directly, through learning of children's speech patterns, or by affecting expectations about child speech), environments where there is a high amount of child-to-child speech will be those most likely to provide toddlers with this experience. Environments such as daycare (whether home-based or center-based) regularly place toddlers in situations where there are more children than adults (e.g., 1 adult per 2–5 children), which is in sharp contrast to home environments where the number of adults typically ranges from 1 to 1.5 per child during the day (Soderstrom et al., 2018). Moreover, Soderstrom et al. (2018) found that the amount of time toddlers spent with other children was far higher in daycare settings than in home environments (even for toddlers at home with siblings). This means that toddlers who spend time in group settings on a regular basis, such as those in daycare, are much more likely to encounter child speech than toddlers who spend the majority of their time at home (with or without siblings).

In addition to the amount of child speech, the variety of speakers toddlers hear may be important. Exposure to more variable input has been shown to improve phonological processing and learning. For example, greater variability has been shown to increase infants' attention to phonetic detail during word learning (Rost & McMurray, 2009, 2010) and their learning of a novel phonotactic pattern (Seidl et al., 2014). Therefore, toddlers who regularly experience group settings (whether in daycare or in some other activity) may be those most likely to develop knowledge of the way that other children speak, regardless of whether or not they have siblings. There are, of course, other features that occur in many (but not all) group settings that could affect children's language knowledge or processing, such as the greater variety of adults encountered and the structured interactions typical of organized activities. However, because we were interested in examining differences in the processing of child speech, we used the amount of time spent with other children in daycare, informal playgroups, and organized activities as our measure of experience, despite these possible co-occurring features.

In the current study, we presented toddlers with visual displays containing two objects (either both familiar or one familiar and one novel) and recorded their looking behavior in response to instructions directing them to look at one of the objects. In Experiment 1, we compared their processing of productions from a child speaker and productions from an adult. In this experiment, words were produced only in their standard form. In Experiment 2, we examined the specificity of toddlers' processing by asking whether they show the same sensitivity to mispronunciations in child speech that they have shown previously for adult speech (e.g., White & Morgan, 2008). Given the research showing that toddlers have difficulty processing the speech of age-matched peers (Cooper et al., 2018; Dodd, 1975) and the lack of studies examining toddlers' processing of slightly older children's speech, we chose a female first-grade student as our child speaker. Although speech from a child this age does not typically contain large deviations from adult target forms, it does have the more subtle deviations described above and clearly has the voice qualities of a child.

Because experience could affect toddlers' processing of child speech, we included the amount of experience with other children as a factor in our analyses. If experience with a diversity of other children sharpens the perception of child speech, toddlers with more experience should show better processing (Experiment 1) and greater phonetic sensitivity (Experiment 2) than those with less experience. If, however, experience with other children primarily leads to more tolerance for deviations in child speech, toddlers with more experience should show equivalent or better processing than those with less experience (Experiment 1) but less phonetic sensitivity to mispronunciations (Experiment 2).

Experiment 1: Recognition of Familiar Words Produced by a Child

In Experiment 1, we compared toddlers' processing of familiar and novel object labels produced by a female child and by a female adult. If child speech is challenging to process, toddlers may be less accurate for words produced by a child, particularly toddlers who have little experience interacting with other children. To determine whether experience is a factor in the processing of child speech, we compared toddlers who spent 8 h or less per week (via parent report) with other children in group settings to those with more experience.

We presented toddlers with three types of trials from each speaker. For the first two types of trials, a label was presented for a familiar target, but the type of distractor object differed (familiar vs. novel). In a situation where both pictured objects have known labels, it may be fairly easy to reject the distractor even if the target label is not pronounced exactly as expected. However, the same is not true when the target object is paired with a novel distractor (see White & Morgan, 2008 for discussion). With a novel distractor, toddlers must decide whether the match between their lexical representation and the label being heard is sufficient, making this a more sensitive test of word recognition. If child speech is harder to process than adult speech, toddlers may be slower or less likely to map familiar labels to familiar objects in the presence of a novel distractor.

It is also possible that there will be differences in processing for the two speakers as a function of the type of label (familiar vs. novel). For this reason, we also included a third type of trial where the novel object was labeled. The task of interpreting a novel label also requires toddlers to evaluate the match between the familiar lexical representation and the label being heard. However, in the case of a novel label, after evaluating (and rejecting) the familiar object as a potential referent, the label must then be mapped to the novel object. If child speech is more difficult to process, toddlers may perform worse on these trials for a child speaker. For example, toddlers could be slower to shift away from the familiar object and toward the novel object within the allotted time window (see Halberda, 2003, for a related finding with younger infants).

Method

Participants. A total of 48 monolingual English-learning toddlers (M = 22.1 months, range = 21.2–23.6) were recruited from the Kitchener/Waterloo region of Ontario, Canada. An additional 3 toddlers were tested but not included due to completing fewer than half of the child speaker trials successfully using the criteria below (n = 2) or parental report indicating that they knew the labels of fewer than half of the familiar objects shown (n = 1).

Half of the participants were classified as High Experience (M = 22.2 months, range = 21.4–23.0) and half were classified as Low Experience (M = 22.0 months, range = 21.2–23.6). Approximately half of the participants in each experience group had older siblings. Experience groups were determined a priori by the average number of hours parents

Table 1. Hours per week participants in Experiment 1 spent with children other than siblings, reported separately for toddlers who did and did not attend daycare

		N	Number of older siblings	Total hours[a] [*M* (range)]	Daycare [*M* (range)]	Other [*M* (range)]
High Experience						
Daycare	First born	10		37.7 (20-48)	35.1 (20-45)	2.6 (0-8)
	Later born	8	1-2	32.4 (18-45)	30.3 (16-45)	2.1 (0-4)
Other	First born	3		15.0 (13-18)	0	15.0 (13-18)
	Later born	3	1	12.0 (10-16)	0	12.0 (10-16)
Low Experience						
	First born	12		2.4 (0-6)	0	2.4 (0-6)
	Later born	12	1-3	1.8 (0-8)	0	1.8 (0-8)

a. Total hours = Daycare + Other

reported that they spent interacting with children other than siblings (see Table 1 for summary). Because it is likely that only a very small number of toddlers have absolutely no regular exposure to children in group settings, we set the cut-off between the experience groups at 8 h per week. Toddlers in the Low Experience group were those who were home with a caregiver (i.e., not in any type of formal childcare setting) and who regularly spent 8 h or less per week interacting with children other than siblings. Those in the High Experience group spent 10 h or more per week with other children outside the home (there were no children whose time in group settings fell between 8 and 10 h). Although most of these toddlers were in daycare, some were home with a caregiver who regularly provided them with activities where there were other children present. Toddlers with these two types of experience are presented separately in Table 1.

Design. There were 24 test trials in total, each with a unique object pair (a list of the familiar objects is presented in Appendix A; sample images are presented in Figure 1). Half of these trials had labels produced by a child and half had labels produced by an adult (blocked, within participant). Of the 12 test trials for each speaker, 8 involved the labeling of familiar objects (4 of these trials had a familiar distractor and 4 had a novel distractor) and 4 involved

Figure 1. Sample image pairs for trials with 2 familiar objects (left), and trials with 1 familiar and 1 novel object (right).

the labeling of novel objects with novel labels. For the novel labels, the two speakers each produced four different labels (paired across speakers such that the labels within a pair were matched for length, syllable structure, and segment type, e.g., *tibble* for the child speaker and *boogle* for the adult speaker). Each pair of novel labels was assigned to a pair of novel objects; the speaker assigned to label each object was counterbalanced across participants.

Two versions of the experiment were generated, each with different target–distractor pairings for both familiar labeled trials (e.g., shoe paired with a bike vs. a novel object) and novel labeled trials (e.g., *tibble* paired with a doll vs. a spoon). Version, speaker order, and objects labeled by each speaker were counterbalanced, producing a total of eight between-participant counterbalancing conditions.

Stimuli. Stimuli were recorded in a sound-attenuated room using a Sennheiser e945 microphone connected to a laptop via a blue icicle USB adaptor. They were recorded into Praat (Boersma & Weenink, 2014). Target words were recorded in one of four sentence frames ("*Where's the ____?*", "*Do you see the ____?*", "*Look! A ____!*", and "*Can you find the ____?*"). The child stimuli were produced by a female first-grade student (7 years 5 months of age). To elicit these productions, the child sat with a female adult in a sound-attenuated room. For each target word, an image of the object appeared on a laptop while the adult produced a labeling sentence in infant-directed speech for the child to repeat. The adult stimuli were recorded in a separate session by the same female adult speaking in an infant-directed speech register. At least three sentence productions per speaker were elicited for each target word; the clearest token of each was selected. Stimuli were later adjusted as necessary by the first author in Praat to ensure that all tokens were of approximately equal perceived intensity (because of the difference in mean pitch noted in Table 2, the mean intensity of the child's productions was set slightly lower than the adult's once equated for perceived intensity). In the testing room, all stimuli were presented at a comfortable listening volume of 65–70 dB.

Table 2 presents the mean acoustic values for the adult and child productions. These measures show that, on average, the child speaker's productions were both shorter and less variable in pitch than the adult's and were also slightly higher in mean pitch. See the online supplementary material for additional information about the productions.

Procedure. Toddlers were tested using the intermodal preferential looking procedure (IPLP) in a sound-attenuated room. Each child sat on the parent's lap while the parent listened to music over circumaural headphones. In front of them was a 42-inch widescreen television and two hidden speakers located at the base of the television; both were connected

Table 2. Mean acoustic values for child and adult speakers

	Child speaker	Adult speaker	Child–Adult difference
Full sentence			
Duration (ms)	1592	1983	-392
F0[a] (Hz)	283	262	21.3
F0 Variation (SD; Hz)	82.4	103.1	-20.7
Target word			
Duration (ms)	664	781	-117
F0 (Hz)	325.8	310.9	14.9
F0 Variation (SD; Hz)	86.2	112.4	-26.1

a. F0 is the acoustical parameter most closely related to perceived pitch. F0 values were calculated using a floor of 100 Hz and a ceiling of 600 Hz for both speakers to a computer in an adjacent room. Participants were monitored over a closed-circuit video feed that was recorded for later offline coding; the camera was centrally located beneath the television and hidden behind a black curtain.

Toddlers were presented with 24 trials in total (12 per speaker). For each speaker, there were four blocks of 3 trials (one trial type per block in pseudorandom order). The four blocks for each speaker were presented in random order. Each trial consisted of images of two objects presented on the left and right sides of the screen. The trial began with the objects shown in silence for 3 s; this was used as the baseline phase. Following the baseline phase, the audio stimulus (e.g., "*Where's the shoe?*") began to play. The images remained on the screen for an additional 5 s from the start of the audio (for a total trial length of 8 s). We defined a priori a 3-s naming phase that commenced 267 ms (8 frames) after the start of the target word (based on the time needed to program an eye movement and previous convention; e.g., Swingley & Aslin, 2000; White & Morgan, 2008). The dependent measure was the change in the proportion of time toddlers spent looking at each object from the baseline phase to the naming phase.

Following the testing session, parents completed a questionnaire on their children's comprehension and production of the experimental items (both familiar and novel) and the amount of time their children interacted with other children (daycare and playgroups/organized activities) on a weekly basis.

Analysis. Looking behavior was coded offline by trained coders blind to condition using customized software at a rate of 30 frames per second (~33.33 ms/frame). For each trial, the proportion of time toddlers spent looking at the labeled object (out of the total time spent looking at the two objects) was calculated for both the baseline and naming phases. For exposition purposes, proportions for the naming phase as a function of time are provided in the supplementary material. However, our dependent measure was a difference score across the entire naming and baseline phase windows (naming minus baseline), which indicates how much toddlers changed their looking to the labeled object after hearing the label. A positive score indicates that they increased their looking to the labeled object following naming, whereas a negative score indicates that they decreased their looking to that object.

For a trial to be included in the analysis, participants needed to look at each of the objects for a minimum of 8 frames (267 ms) during the baseline phase. This criterion resulted in 16.2% of trials being discarded (High Experience: 15.6%; Low Experience: 16.8%). Participants also needed to attend to the objects for a minimum of 1 s total during each phase. This criterion resulted in 1% of trials being discarded (High Experience: 0.9%; Low Experience: 1%).

Data were analyzed using linear mixed effects regression (Baayen, Davidson, & Bates, 2008) with R's lme4 package (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2018). Fixed effects of experience (high vs. low), speaker (adult vs. child),[1] and condition (familiar labeled/familiar distractor vs. familiar labeled/novel distractor vs. novel labeled) were contrast-coded. The first condition contrast compared the two conditions where the familiar object was labeled and the second condition contrast compared these first two conditions with the final condition where the novel object was labeled. The first condition contrast tests the effect of competitor type on familiar word recognition; the second tests the effect of familiar versus novel object labeling. The lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017), which uses Satterthwaite's method for approximating degrees of freedom, was used to assess the significance of intercepts as well as contrasts of simple

[1] Although toddlers in the High Experience group showed an overall increase in looking for the second speaker, whereas those in the Low Experience group did not, test order did not interact with condition for either experience group and, therefore, was not included in our main analyses.

models with no more than one fixed effect (Luke, 2017). Random effects error variance was used as an estimate of σ^2 to calculate Cohen's *d* for model estimates (Brysbaert & Stevens, 2018; Westfall, Kenny, & Judd, 2014).

Results

Baseline looking. We first assessed looking during the baseline phase against chance (0.5). To do this, baseline looking was centered around 0 by subtracting 0.5 from each value, thereby making the intercept of the following models equivalent to the deviation from chance. Linear mixed effects regression models were generated separately for each type of display (two familiar objects or one familiar object and one novel object), with experience entered as a fixed effect along with random intercepts for participant and item (i.e., the specific object being labeled). Because baseline performance should be unaffected by speaker, condition, and test version,[2] by-participant and by-item random slopes were not included.

Considering first trials with two familiar objects, model estimates indicate that, overall, toddlers spent equivalent amounts of time on each object (β_0 = 0.034, β_0 *SE* = 0.021, $t(14.8)$ = 1.613, p = .128, d = 0.17), and this did not differ across experience groups (β_1 = 0.003, β_1 *SE* = 0.022, $t(286.4)$ = 0.140, p = .889, d = 0.02). For trials with one familiar object and one novel object, baseline looking was assessed as looks to the *familiar* object. Model estimates indicate that, overall, toddlers spent significantly more time on the familiar object (β_0 = 0.071, β_0 *SE* = 0.015), $t(22.2)$ = 4.618, p < .001, d = 0.36). This is in keeping with other work showing a familiarity bias during baseline with this kind of display (e.g., White & Morgan, 2008). However, the Low Experience group spent significantly less time on the familiar object than

[2] Recall that version refers to the two different target–distractor pairings generated for each target object (see "Design" section above).

the High Experience group (β_1 = -0.040, β_1 *SE* = 0.014, *t*(629.1) = -2.820, *p* = .005, *d* = -0.21).

Word recognition. Our main analyses involved the naming-baseline difference scores. Condition, Speaker, and Experience, as well as their interactions, were entered into the model as fixed effects. The maximum random effects structure that allowed for convergence included random intercepts for participant and item as well as by-participant random slopes for speaker and by-item random slopes for version. To assess whether the inclusion of each effect (main effect and interaction) significantly added to the model, maximum likelihood models were compared with and without the effect of interest, with all other possible effects included. These analyses revealed only a significant Condition x Experience interaction, $\chi^2(2) = 14.193$, $p < .001$, with model estimates indicating that experience had an effect on performance for the contrast testing the type of distractor object (β = 0.106, β *SE* = 0.049, *t* = 2.152, *d* = 0.33), as well as the contrast testing the type of target object (β = -0.128, β *SE* = 0.042, *t* = -3.056, *d* = -0.40). No other effects were significant (χ^2s ≤ 2.15, *ps* ≥ .347), including all of those involving speaker.[3]

To further examine the Condition x Experience interaction (Figure 2), separate models were generated for each experience group with condition entered as a fixed effect. The

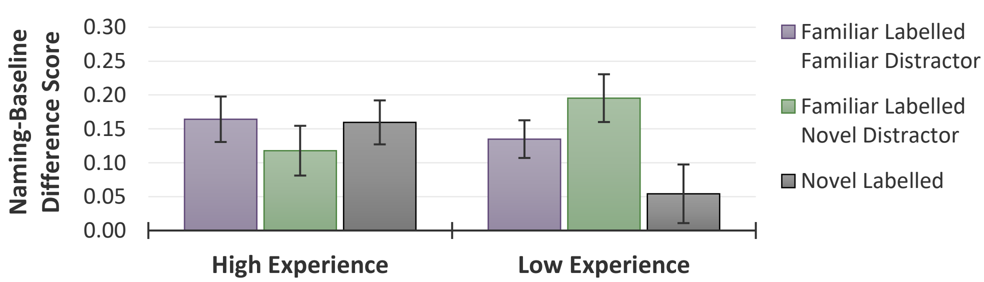


Figure 2. Estimated change in looking to the *labeled* object for each condition as a function of experience group in Experiment 1. A positive score indicates an increase in looking at the labeled object following naming. Error bars represent standard errors.

[3] Model comparisons show that Birth Order had no effect on the Condition x Experience interaction (χ^2s ≤ 2.00, *ps* ≥ .176).

maximum random effects structure that allowed for convergence included random intercepts for participant and item as well as by-participant random slopes for speaker for the High Experience group and by-item random slopes for version for the Low Experience group. In addition, each condition's estimated difference from baseline (0 or no change; see Table 3) was assessed using intercept-only models (no fixed effects and only random intercepts for participant and item were included).

First, in examining the effect of distractor type, no difference between familiar and novel distractors was found for either the High Experience group (β_1 = -0.052, β_1 *SE* = 0.034, $t(428.8)$ = -1.538, p = .125, d = -0.16), or the Low Experience group (β_1 = 0.059, β_1 *SE* = 0.043, $t(22.7)$ = 1.377, p = .182, d = 0.17). However, the effects are in opposing directions for the two groups, thereby explaining the interaction for this contrast. Assessment of the estimated differences from baseline (0 or no change) show that toddlers in both experience groups significantly increased their looking to the target object for familiar labels regardless of speaker or distractor, ps ≤ .025, ds ≥ 0.15. Therefore, whether the distractor object was a familiar versus novel object did not affect familiar label processing for either experience group.

For the type of target object (familiar vs. novel), we found that, whereas toddlers in the High Experience group looked to the target object equally for familiar and novel labels (β_2 = 0.019, β_2 *SE* = 0.048, $t(22.3)$ = 0.398, p = .694, d = 0.06), toddlers in the Low Experience group

Table 3. Estimated proportion changes (SE) in looking to the labeled object for each speaker as a function of condition and experience group

		Familiar labeled, Familiar distractor	Familiar labeled, Novel distractor	Novel labeled
High Experience	Adult speaker	0.182 (0.046)	0.092 (0.037)	0.189 (0.012)
	Child speaker	0.153 (0.044)	0.142 (0.053)	0.132 (0.035)
Low Experience	Adult speaker	0.105 (0.040)	0.173 (0.037)	0.052 (0.028)
	Child speaker	0.149 (0.047)	0.214 (0.045)	0.051 (0.047)

increased their looking to the target significantly less for novel labels than for familiar ones (β_2 = -0.108, β_2 *SE* = 0.039, *t*(22) = -2.74, *p* = .012, *d* = -0.32). Assessments against baseline show that when the novel object was labeled, toddlers in the High Experience group increased their looking to the target object for both speakers, *p*s ≤ .009, *d*s ≥ 0.23, whereas toddlers in the Low Experience group did not change their looking from baseline for either speaker, *p*s ≥ .318, *d*s ≤ 0.08. Supporting this difference across experience groups, models assessing effects of experience (with random intercepts for participant and item) revealed a significant difference across experience groups for novel labels (β_1 = -0.106, β_1 *SE* = 0.043, *t*(42.1) = -2.482, *p* = .017, *d* = -0.31), but not familiar labels (β_1 = 0.028, β_1 *SE* = 0.029, *t*(45) = 0.928, *p* = .358, *d* = 0.09).

Discussion

When familiar objects were labeled, toddlers responded similarly regardless of the type of distractor or their experience with other children. However, this was not the case with novel labels. On these trials, only toddlers with higher levels of experience with other children showed a disambiguation response and increased their looking to the novel object. Interestingly, these patterns were true regardless of speaker. Therefore, our task did not reveal any differences in the processing of speech produced by an adult and by a 7-year-old child.

Although these results suggest that experience influences the mapping of novel words, one concern is that the effect of experience for novel label trials may have been driven by differences in baseline preference across the groups. In other words, perhaps the toddlers in the High Experience group were more likely to shift to the novel object during the naming phase because of their overall stronger preference for the familiar object during baseline. However, we have at least three arguments against this concern. First, looking patterns during the naming phase alone mirror the reported difference scores for both speakers. Namely, although both experience groups looked to the familiar object at above chance levels after hearing a familiar label (*p*s ≤ .009, *d*s ≥ 0.44), only the High Experience group trended toward looking at the novel object after hearing a novel label (*p*s ≤ .059, *d*s ≥ 0.29 for the High Experience group vs. *p*s ≥ .615, *d*s ≤ 0.09 for the Low Experience group). Second, if baseline preference determined the degree of naming preference, we would expect a negative correlation between the amount of time spent looking at the familiar object during the baseline and naming phases. This was not the case. Third, baseline preferences on adult speaker trials differed more as a function of experience than baseline preferences on child speaker trials (*d* = -0.33 vs. -0.11), yet toddlers changed their looking equivalently for both speakers. Therefore, differences in baseline preference do not explain the effect of experience for novel labels.

In summary, we found that toddlers, regardless of experience with other children, were equally likely to look to the familiar object when familiar labels were produced by a child speaker vs. an adult speaker. In other words, toddlers exhibited no difficulty processing the speech of an early school-aged child in this task. However, all familiar labels were pronounced correctly. This means that toddlers may have been able to successfully deduce the target for the child speaker through the use of a more general matching strategy that did not require them to pay close attention to the phonetic detail. Next, we turned to an approach that has been fruitful in testing toddlers' sensitivity to phonetic detail during processing – that of mispronunciation detection.

Experiment 2: Detection of Mispronunciations in Child Speech

Previous work has demonstrated that, for familiar words produced by adults, toddlers are very sensitive to phonetic changes, detecting even slight mispronunciations of those words (e.g., Swingley & Aslin, 2000). For example, toddlers look to an image of a baby less, and more slowly, when it is labeled as *vaby* compared to when it is labeled correctly. Not only do toddlers show processing costs for mispronunciations, they also show graded sensitivity to the degree of mispronunciation (Mani & Plunkett, 2011; Ren & Morgan, 2011; Tamási et al., 2017; White & Morgan, 2008). In looking paradigms, toddlers progressively decrease their looking to the target object following labeling as the degree of mispronunciation increases (White & Morgan, 2008). For example, there may be a relatively small cost associated with a one-feature phonetic change, but there may be a larger cost associated with a two-feature change and an even larger cost associated with a three-feature change. This graded sensitivity has also been demonstrated using pupillometry, where the larger the degree of mispronunciation, the greater the change in pupil diameter (Tamási et al., 2017).

In Experiment 2, we used a mispronunciation procedure to tap toddlers' sensitivity to phonetic mismatch in speech produced by a child. Labels for familiar objects were pronounced either correctly or with an onset mispronunciation of one to three phonetic features. We envisioned at least two possible outcomes. The first was that toddlers would show graded sensitivity to the degree of mispronunciation, with a penalty for even single-feature mispronunciations (the same pattern of response for a child speaker as toddlers have shown in previous research for an adult speaker). This would indicate high sensitivity to phonetic detail in child speech. A second possible outcome was that at least a subset of toddlers would accept anything close to a correct label from a child speaker. This outcome could arise for two reasons. If it were due to difficulty resolving subtle differences (i.e., reduced sensitivity to phonetic detail) in child speech, we would expect to see reduced mispronunciation penalties in the group of toddlers with *less* experience with child speech. If, on the other hand, it were due to a greater tolerance for child mispronunciations, we would

Table 4. Hours per week participants in Experiment 2 spent with children other than siblings, reported separately for those who did and did not attend daycare

		N	Number of older siblings	Total hours[a] [*M* (range)]	Daycare [*M* (range)]	Other [*M* (range)]
High Experience						
Daycare	First born	13		34.6 (18-50)	32.2 (16-45)	2.4 (0-8)
	Later born	5	1	27.4 (11-37)	26.0 (10-35)	1.4 (0-3)
Other	First born	2		12.5 (10-15)	0	12.5 (10-15)
Low Experience						
	First born	15		3.0 (0-5)	0	3.0 (0-5)
	Later born	5	1-2	3.6 (0-7)	0	3.6 (0-7)

a. Total hours = Daycare + Other

expect to see reduced mispronunciation penalties for toddlers with *more* experience with child speech.

Method

Participants. A total of 40 monolingual English-learning toddlers (*M* = 21.7 months, range = 20.9–23.2) participated in this experiment. An additional 12 toddlers were tested but not included due to technical issues with equipment (*n* = 3), fussiness/crying (*n* = 4), completing fewer than half of the trials successfully using the criteria below (*n* = 2), parental interference (*n* = 1), not successfully completing at least two correct label trials (*n* = 1), or performance on correct label trials more than 3.5 standard deviations below the mean (*n* = 1).

Half of the participants were classified as High Experience (*M* = 21.6 months, range = 20.9–22.6) and half were classified as Low Experience (*M* = 21.8 months, range = 21.0–23.2). Experience groups were determined in the same way as in Experiment 1. This time a quarter of the participants in each experience group had older siblings (see Table 4 for summary).

Design. There were 20 test trials in total, each with a unique familiar object–novel object pair (a list of the familiar objects is presented in Appendix A). Of these 20 trials, the familiar object was labeled correctly on 4 trials, the novel object was labeled with a novel word (e.g., *tibble*) on 4 trials, and the familiar object was labeled with one of three types of mispronunciations (4 trials for each type) on the remaining 12 trials. These mispronunciations consisted of onset consonant changes of one to three features that resulted in either a nonword or a word that toddlers are unlikely to know (see Table A2 for the full set of mispronunciations).

Each participant heard one type of pronunciation (correct, one-feature change, two-feature change, or three-feature change) for each familiar target object. The assignment of pronunciations to familiar objects was counterbalanced across participants. Novel label trials used different familiar object–novel object pairs and were the same across all participants.

There were four blocks of 5 trials (one type of trial per block in pseudorandom order). The blocks were presented in random order. An additional 4-s trial presented the image of a young girl before each block of trials.

Stimuli. Correctly pronounced familiar labels and novel labels were the same as those used in Experiment 1. Mispronounced labels were recorded by the child during the same recording session and were elicited in the same way. Stimuli were presented to participants at a comfortable listening volume of 65–70 dB. Acoustic measures for the stimuli are

Table 5. Mean acoustic values for the child speaker for each condition

	Correct	1-Feature Change	2-Feature Change	3-Feature Change	Novel[a]
Full Sentence					
Duration (ms)	1582	1665	1633	1613	1626
F0[b] (Hz)	281.2	279.0	273.0	277.8	289.0
F0 Variation (SD; Hz)	82.8	76.8	72.6	76.9	76.2
Target Word					
Duration (ms)	650	675	658	639	606
F0 (Hz)	326.3	322.7	302.7	321.3	301.4
F0 Variation (SD; Hz)	83.9	90.7	84.3	80.3	98.4

a. n = 4 novel labels compared to n = 16 for each of the other conditions
b. F0 is the acoustical parameter most closely related to perceived pitch. F0 values were calculated using a floor of 100 Hz and a ceiling of 600 Hz

presented in Table 5, and show that the child speaker's productions of the mispronounced labels did not differ systematically from their familiar counterparts.

Procedure. The procedure was the same as in Experiment 1.

Analysis. As in Experiment 1, looking behavior was coded offline by trained coders blind to condition using customized software at a rate of 30 frames per second (~33.33 ms/frame). This time, however, the difference score (our dependent measure) was computed based on the proportion of time toddlers spent looking at the *familiar* object. Once again, for exposition purposes, proportions for the naming phase as a function of time are provided in the supplementary material. The same criteria for trial exclusion as in Experiment 1 were used. This resulted in 17.3% of trials being discarded for participants not attending to both objects during baseline (High Experience: 18.0%; Low Experience: 16.5%) and 1.0% of trials being discarded for not being on-task for at least 1 s in each of the baseline and naming phases (High Experience: 1.0%; Low Experience: 1.0%).

As in Experiment 1, data were analyzed using linear mixed effects regression. The fixed effect of experience (high vs. low) was contrast-coded, whereas the fixed effect of condition (correct vs. one-feature vs. two-feature vs. three-feature vs. novel) remained treatment-coded to enable a comparison of correct against each of the mispronunciation conditions.

Results

Baseline looking. We first assessed baseline preferences against chance (.50). To do this, baseline looking was centered around 0. Experience was entered as a fixed effect into a linear mixed effects regression model along with random intercepts for participant and item (i.e., the specific object being labeled). Model estimates indicate that, just like in Experiment 1, toddlers spent significantly more time looking at the familiar object (β_0 = 0.097, β_0 SE = 0.014), $t(19.9) = 7.162$, $p < .001$, $d = 0.53$), and toddlers in the Low Experience group spent marginally less time on the familiar object than those in the High Experience group (β_1 = -0.031, β_1 *SE* = 0.015), $t(34.7) = -2.013$, $p = .052$, $d = -0.17$).

Word recognition. As in Experiment 1, our main analyses involve the naming–baseline difference scores. The maximum random effects structure that allowed for convergence for each of the following analyses included only random intercepts for participant and item. We first considered only trials in which the familiar object was labeled correctly, with experience entered into the model as a fixed factor. This analysis showed that toddlers recognized these words, increasing their looking to the familiar object from baseline (β_0 = 0.161, β_0 *SE* = 0.030, $t(14.7) = 5.317$, $p < .001$, $d = 0.53$), with no difference across experience groups (β_1 = -0.009, β_1 *SE* = 0.063, $t(35.7) = 0.951$, $p = .857$, $d = -0.03$). Evaluation of the estimated difference from baseline (0 or no change) via intercept-only models showed that both the High Experience group (β_0 = 0.169, β_0 *SE* = 0.035) and the Low Experience group (β_0 = 0.152, β_0 *SE* = 0.049) significantly increased their looking to the familiar object (*p*s ≤ .007, *d*s ≥ 0.48).

We next assessed whether there was an effect of pronunciation type across the four conditions in which the familiar object was labeled (correct vs. one-feature vs. two-feature vs. three-feature). Condition, experience, and their interaction were entered into the model as fixed effects. Maximum likelihood model comparisons were performed to assess whether the inclusion of each effect significantly added to the model. This analysis revealed only a significant main effect of condition, $\chi^2(3) = 41.42$, $p < .001$, with model estimates indicating that toddlers were less likely to increase their looking to a familiar object for mispronounced labels than for correct ones and that this was true for all degrees of change (one-feature: β_1 = -0.088, β_1 *SE* = 0.040, $t = -2.186$, $d = -0.27$; two-feature: β_2 = -0.149, β_2 *SE* = 0.040, $t = -3.716$, $d = -0.45$; three-feature: β_3 = -0.255, β_3 *SE* = 0.040, $t = -6.349$, $d = -0.77$). No other effects reached significance (χ^2s ≤ 1.29, *p*s ≥ .455).

Given our hypotheses about the role of child experience, we conducted exploratory analyses to examine the effects of experience despite the lack of interaction involving this factor. To do so, models of the effect of condition were generated separately for each experience group (Figure 3). For the High Experience group, the model showed that all three types of mispronunciations significantly differed from the correct condition (one-feature: β_1 = -0.114, β_1 *SE* = 0.057, *p* = .047, *d* = -0.35; two-feature: β_2 = -0.154, β_2 *SE* = 0.056, *p* = .007, *d* = -0.47; three-feature: β_3 = -0.297, β_3 *SE* = 0.057, *p* < .001, *d* = -0.90). Evaluation of each condition's estimated difference from baseline showed that toddlers did not change their looking for either a one-feature change (β_0 = 0.051, β_0 *SE* = 0.044, *p* = .258, *d* = 0.14) or a two-feature change (β_0 = 0.012, β_0 *SE* = 0.043, *p* = .777, *d* < 0.01), but they significantly decreased their looking to the familiar object for a three-feature change (β_0 = -0.124, β_0 *SE* = 0.055, *p* = .039, *d* = -0.35). In contrast, the model for the Low Experience group showed that the smallest change was not treated differently from the correct condition (one-feature: β_1 = -0.058, β_1 *SE* = 0.056, *p* = .305, *d* = -0.17), but larger changes were (two-feature: β_2 = -0.142, β_2 *SE* = 0.057, *p* = .012, *d* = -0.43; three-feature: β_3 = -0.211, β_3 *SE* = 0.057, *p* < .001, *d* = -0.63). Estimated differences from baseline showed that toddlers significantly increased their looking to the familiar object for a one-feature change (β_0 = 0.094, β_0 *SE* = 0.040, *p* = .034, *d* = 0.29), and showed no difference from baseline for a two-feature change (β_0 = 0.012, β_0 *SE* = 0.057, *p* = .831, *d* = 0.03) or a three-feature change (β_0 = -0.06, β_0 *SE* = 0.05, *p* = .658, *d* = -0.16). Despite these somewhat different patterns of significance for the two groups, models that include only the three types of mispronunciations coded to assess a polynomial fit reveal significant linear trends for both the High Experience group (β_1 = -0.128, β_1 *SE* = 0.043, $t(184.1)$ = -3.016, *p* = .003) and the Low Experience group (β_1 = -0.109, β_1 *SE* = 0.040, $t(171.8)$ = -2.749, *p* = .007). Together, these analyses indicate that, although toddlers in the Low Experience group showed somewhat reduced mispronunciation penalties compared with toddlers in the High Experience group, both groups were sensitive to the degree of mispronunciation in child speech.

Finally, we looked at performance on novel label trials with experience entered into the model as a fixed factor. Overall, toddlers did not significantly change their looking to the familiar object from baseline (β_0 = -0.065, β_0 *SE* = 0.037, *t*(3.3) = -1.827, *p* = .156, *d* = -0.19), with no significant difference across experience groups (β_1 = 0.060, β_1 *SE* = 0.064, *t*(35.7) = 0.951, *p* = .348, *d* = 0.18). However, as can be seen in Figure 3, whereas the High Experience group trended toward the novel object (β_0 = -0.096, β_0 *SE* = 0.056, *t*(6.4) = -1.75, *p* = .134, *d* = -0.29), the Low Experience group did not change their looking from baseline (β_0 = -0.038, β_0 *SE* = 0.067, *t*(2.9) = -0.492, *p* = .658, *d* = -0.09). This pattern is similar to that in Experiment 1, where toddlers in the High Experience group looked to the target object equivalently for familiar and novel labels, but toddlers in the Low Experience group looked to the target object significantly less for novel labels than for familiar ones. This suggests that, as in Experiment 1, it is in the processing of novel labels that effects of experience appear. To explore this possibility, we assessed the similarity of these patterns across the two

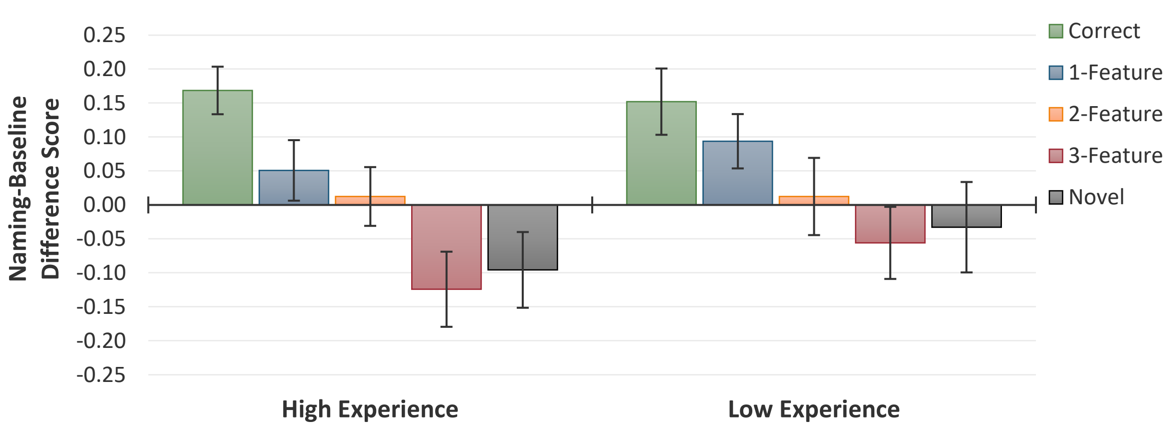


Figure 3. Estimated change in looking to the *familiar* object for each condition as a function of experience group in Experiment 2. A positive score indicates an increase in looking at the *familiar* object following naming. Error bars represent standard errors.

experiments directly.

Novel label processing across experiments. To compare performance on familiar and novel label trials across experiments, we compared responses for the two trial types that appeared in both experiments (for Experiment 1, only trials with the child speaker were included) – trials involving familiar–novel object pairs with either correctly pronounced familiar labels or novel labels. To do this, we first recoded the novel label data for Experiment 2 so that they represented *looks to the labeled object* (as in Experiment 1). We then pooled the data for these two trial types across experiments, resulting in 44 toddlers in each experience group. Condition (familiar vs. novel), experience (high vs. low), and Experiment (1 vs. 2) were entered into the model as contrast-coded fixed effects along with their interactions. The maximum random effects structure that allowed for convergence consisted of random intercepts for participant and item[4], as well as a by-item random slope for experiment. Maximum likelihood model comparisons revealed a main effect of condition, $\chi^2(1) = 4.59$, $p = .032$, as well as a Condition x Experience interaction, $\chi^2(1) = 4.75$, $p = .029$. The model showed that overall toddlers increased their looking to the target object less for a novel label than for a familiar one ($\beta = -0.092$, β *SE* = 0.041, $t = -2.272$, $d = -0.28$), and this difference was more pronounced for the Low Experience group than the High Experience group ($\beta = -0.99$, β *SE* = 0.051, $t = -1.930$, $d = -0.31$). No other effects reached significance, χ^2s ≤ 0.93, *p*s $\geq .335$ (Figure 4).

To assess the Condition x Experience interaction (Figure 4), separate models for each experience group were generated with condition as a fixed effect along with random intercepts for participant and item (the maximum random effects structure). These analyses

[4] Due to the use of different novel object-label pairings across participants, for this analysis item represented the label used rather than the object being labelled.

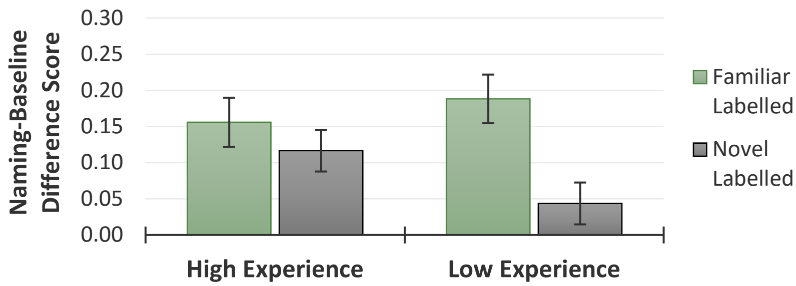


Figure 4. Estimated change in looking to the *labeled* object for familiar labels in the presence of a novel distractor and novel labels as a function of experience group for the child speaker pooled across Experiments 1 and 2. A positive score indicates an increase in looking at the *labeled* object following naming. Error bars represent standard errors.

showed that toddlers in the High Experience group increased their looking to the target object similarly for familiar and novel labels (β_1 = -0.040, β_1 *SE* = 0.046, *t*(4.8) = -0.861, *p* = .430, *d* = -0.13), whereas toddlers in the Low Experience group increased their looking to the target object for novel labels significantly less than they did for familiar labels (β_1 = -0.143, β_1 *SE* = 0.041, *t*(6.1) = -3.448, *p* = .013, *d* = -0.43). Intercept-only mixed effects models for each condition show that toddlers in both the High Experience and Low Experience groups significantly increased their looking to the target object for familiar labels, (β_0s ≥ 0.156, *ps* < .001, *d* ≥ 0.54), but only those in the High Experience group did so for novel labels (β_0 = 0.117, β_0 *SE* = 0.023, *t*(38.8) = 4.045, *p* < .001, *d* = 0.38). Toddlers in the Low Experience group did not change their looking from baseline for novel labels (β_0 = 0.044, β_0 *SE* = 0.029, *t*(149) = 1.505, *p* = .134, *d* = 0.12). This indicates that, across experiments, all toddlers recognized the familiar objects, but only toddlers in the High Experience group showed a disambiguation response for the novel labels.

Discussion

When object labels were pronounced correctly, toddlers successfully mapped them to the appropriate referents. When these same familiar object labels were mispronounced, toddlers' looking behavior differed from the correct pronunciations, showing that overall they were sensitive to each level of change. However, subtle differences arose in the treatment of one-feature and three-feature changes across experience groups. Toddlers in the High Experience group demonstrated somewhat greater sensitivity to phonetic differences, showing a larger penalty for both one-feature and three-feature mispronunciations than toddlers in the Low Experience group. Although the difference across groups was not significant, this pattern suggests that toddlers with limited experience with other children may have some difficulty resolving subtle differences in child speech. That said, regardless of experience, toddlers showed graded sensitivity to the degree of change, with progressively larger mispronunciation penalties for one-feature, two-feature, and three-feature changes, a pattern that is similar to the one previously found for toddlers' processing of adult speech (White & Morgan, 2008). Thus, despite some differences between the two experience groups, toddlers were overall very sensitive to the phonetic content of the child speech, suggesting that high levels of child experience are not required to demonstrate this sensitivity. Finally, this experiment also provides additional evidence that toddlers with varying levels of child experience differ in their treatment of novel labels. We return to this topic in the General Discussion.

Chapter Discussion: Experiments 1 and 2

Across two experiments, we have shown that toddlers process familiar labels from a 7-year-old speaker in our task as accurately, and with as much sensitivity, as they process labels from an adult. For the most part, this sensitivity was unaffected by toddlers' experience with other children. However, only toddlers with experience in interacting regularly with a variety of other children showed a disambiguation response for novel labels.

We predicted that the effects of increased experience might manifest in one of two ways: first, as increased accuracy in processing child speech (i.e., higher looking to the target in

Experiment 1 and higher sensitivity to phonetic changes in Experiment 2) or, second, as increased tolerance for deviations in child speech (in Experiment 2). Our results were not strongly in line with either of these predictions. Instead, we found that toddlers were equally accurate in processing familiar labels regardless of their amount of experience with child speech, although there was a hint that toddlers with less experience showed reduced sensitivity to mispronunciations in Experiment 2 (consistent with some effect of experience on processing the fine details of child speech).

It may be that our child speaker was too old (and her productions too mature) for us to have seen significant effects on processing. In two previous studies (Cooper et al., 2018; Dodd, 1975), 2- to 4- year-old children were less accurate in processing their own speech, as well as the speech of another child, compared with that of an adult. One obvious difference between studies is the age of the child speaker (2–4 years vs. 7 years). Together, these findings suggest that toddlers are able to cope with the more subtle deviations that older children's speech may exhibit and that it is the larger distortions (e.g., substitutions, deletions) that impair comprehension. Nonetheless, we cannot rule out the possibility that other measures (e.g., pupil dilation) or more challenging listening conditions (e.g., with background noise) might reveal differences in the processing of speech from adults and older children. Future work could explore speech from a greater range of speaker ages, and using a greater range of tasks, to determine at what age children's speech is processed in a mature manner and whether experience with speech from other children is more important for processing younger children's speech.

Our findings suggest that toddlers are quite adept at processing the speech of a child speaker, but our study leaves open the question of what assumptions toddlers made about the age of the speaker. Adults are extremely sensitive to the acoustic markers of speaker age

(Amir, Engel, Shabtai, & Amir, 2012; Cerrato, Falcone, & Paoloni, 2000) and would have recognized that the child speaker was young based on the acoustic properties of her voice (e.g., higher mean pitch). However, it is unclear at what age this ability develops. Because we were purely interested in toddlers' ability to process the speech itself, our task did not require that they be aware of the speakers' ages. Future work could explore when children begin to map acoustic properties to speaker age.

Although we did not find a difference in accuracy for the child and adult speakers, our results showed that toddlers who had little exposure to other children did not change their looking from baseline after hearing a novel label and that, unexpectedly, this was true for both the child and adult speakers. In other words, this group did not show the expected disambiguation response that is typically seen in monolingual toddlers as young as 17 months when a novel label is presented by an adult (Byers-Heinlein & Werker, 2009; Halberda, 2003; White & Morgan, 2008). Given that this was true for both speakers, it seems unlikely that this is due to difficulties in processing child speech in particular. Why is it that only toddlers whose experience came from a variety of child speakers, such as occurs in a daycare setting, successfully mapped novel labels to novel objects? Although we focused on their amount of experience with other children, toddlers who are in group settings are also likely to hear a greater variety of adult speakers (although there are certainly some exceptions to this, e.g., homebased daycare settings with only one adult). Therefore, it is possible that there is an influence of speaker variability more generally. Previous work has demonstrated that speaker variability introduced during the learning process affects attention to phonetic detail and the generalizability of word representations (e.g., Rost & McMurray, 2009; Singh, 2008). If speaker variability played a role here, it would suggest that hearing speech from a greater variety of speakers (child and/or adult) affects the efficiency or robustness of *subsequent* word learning with new speakers.

Another possibility is that exposure to multiple speakers in the environment has effects on vocabulary size (perhaps via the influence of variability or as a result of having a wider variety of interactions in which different topics or objects are discussed) and that this in turn affects novel label processing. Indeed, the connection between number of speakers and vocabulary size has been demonstrated for bilingual children (Place & Hoff, 2011), as has the connection between vocabulary size and the use of a disambiguation strategy (Bion, Borovsky, & Fernald, 2013). Together, these studies are consistent with the possibility that toddlers with language exposure from a wider variety of individuals (all else being equal) may have larger vocabularies and that these toddlers with larger vocabularies will be more likely to disambiguate novel labels. However, this connection remains speculative at this point.

Although we did not test vocabulary levels directly, one aspect of our findings that is potentially relevant is the difference in baseline familiarity bias between the High Experience and Low Experience groups for trials with a familiar–novel object pair. In both Experiment 1 and Experiment 2, toddlers with more experience with other children showed a stronger baseline preference for the familiar object. This stronger preference could indicate that toddlers in the High Experience group were more familiar with the familiar object labels (Schafer, Plunkett, & Harris, 1999). Previous work has shown that children's degree of knowledge about specific familiar labels predicts their disambiguation response when a novel label is presented (Grassmann, Schulze, & Tomasello, 2015). Therefore, it is possible that weaker knowledge of individual familiar labels was responsible for the difficulty toddlers in the Low Experience group had with novel labels. It may also be that knowledge of individual familiar labels is correlated with vocabulary size, consistent with a link between processing efficiency and vocabulary size (Fernald et al., 2006). To more directly investigate the potential role of vocabulary in explaining our findings, we conducted a follow-up with a separate sample of same-aged children drawn from the same population (N = 47; 24 female, 20 male, and 3 undisclosed). In this follow-up, we administered parent questionnaires and the MacArthur Communicative Development Inventories short form (Fenson et al., 2000) to examine the relationship between vocabulary size and experience in group settings. However, inconsistent with a role for vocabulary size, we found no relationship between vocabulary size and the amount of time toddlers spent with other children outside the home (r = -0.08). In other words, toddlers' vocabulary size was similar regardless of whether they attended daycare, engaged in regular playgroups, or spent the majority of their time at home. In addition, maternal education did not differ across these three groups, with more than 80% of mothers in each group having 3 or more years of postsecondary education. Although not conclusive, these survey results (drawn from the same population) suggest that neither the quality of maternal input (as approximated by maternal education) nor toddlers' vocabulary size was responsible for the difference in novel label processing in the current study.

The lack of a relationship between care setting and vocabulary in our follow-up survey is not entirely surprising given that previous literature has not found consistent effects of daycare on language outcomes. For example, Booth, Clarke-Stewart, Vandell, McCartney, and Owen (2002) observed no differences in 15-month-olds' vocabulary scores as a function of the children's daycare status, whereas Laing and Bergelson (2018) found that 17-month-olds with a combination of home care and daycare had larger vocabularies than toddlers with either home care or daycare alone. But in a large-scale NICHD ECCRN (2000) study, 3-year-olds in high-quality daycare did have larger vocabularies than children in other care situations. Thus, overall, it appears to be the quality of care that is most important (Burchinal, Roberts, Nabors, & Bryant, 1996; McCartney, 1984), although there is considerable debate about how quality of care should be assessed (see Falenchuk, Perlman, McMullen, Fletcher, & Shah, 2017; Perlman et al., 2016, 2017 for meta-analyses of factors such as the education of staff, child–staff ratios, and child–staff interaction quality). Although our discussion so far has focused on how the number of speakers and vocabulary size might contribute to novel word mapping, it is also possible that other factors differed across our experience groups. One possibility is that environments such as daycare settings provide more opportunities for structured word learning play, and that this in turn affected performance on our novel label trials. However, there are a wide variety of daycare settings, not all of which provide these very organized and deliberate kinds of learning experiences to the same degree. In addition, not all of the toddlers in our High Experience group had experience with formal childcare settings; some had experience with other children through informal playgroups and community-based activities such as swimming and gymnastics. Therefore, it remains possible that something about interacting with other children (or a variety of different individuals) may provide toddlers with the kinds of experience that boost novel label processing. Because this was not the primary goal of the current study, future work will be needed to examine more systematically this intriguing relationship among novel word processing, type of experience with other children, and other experiential factors.

In conclusion, toddlers are exposed not only to the speech of adults, but also to the speech of other children. The current study is the first demonstration that toddlers show considerable sensitivity to phonetic detail in the speech of a child, processing it as well as that of an adult. This was true (although to a somewhat lesser extent) even for toddlers with less experience hearing other children. Given this sensitivity, speech from children of this age could be useful input for young language learners. A full picture of the impact that this has on toddlers' language development will require additional research, especially because some studies of toddlers in other cultures (e.g., the Mayan Yucatec) indicate that their vocabulary size is best predicted by the quantity of speech directed at them from *adults* and not from the children with whom they spend most of their time (Shneidman & Goldin-Meadow, 2012). Finally, our findings also demonstrate that toddlers with more versus less exposure to speech from other children show a difference in their treatment of novel labels even when they are produced by adults. Whether this is truly due to the number or variety of other child speakers or to other correlated aspects of toddlers' language or broader environment is unclear. These are intriguing questions for future research.

Chapter Three: Sensitivity to Frequency of Occurrence in Child Speech

The following chapter has been published as:

Bernier, D. E., & White, K. S. (2019b). Toddlers process common and infrequent childhood mispronunciations differently for child and adult speakers. *Journal of Speech, Language, and Hearing Research*, *62*, 4137-4149. https://doi.org/10.1044/2019_JSLHR-H-18-0465

Toddlers' environments are filled with people of different ages who speak in different ways – from mature adult speakers to other young language learners. Although adult speakers show some variability in their productions of speech sounds, very young children show much higher variability and are more likely to make phonological errors, such as substituting one sound for another. However, these substitutions are not random; some, such as saying *fumb* for thumb, are very common for young children to make, whereas others, such as saying *pumb* for thumb, are rarely observed (Smit, 1993). In the current article, we investigate toddlers' processing of these different types of substitutions.

The majority of young children's phonological errors (whether substitutions or other types of errors, such as phoneme deletions) decrease by the age of 4 years, and they are largely absent after the age of 6 years (Dodd et al., 2003). In keeping with this gradual improvement with age, the speech of very young children can be difficult for naïve adult listeners to understand (Flipsen, 2006; Hodson & Paden, 1981), whereas the productions of slightly older children are (on average) quite intelligible. However, an adult's experience with young children can influence intelligibility (Munson et al., 2012). The fact that the speech of young children is difficult to process for even mature language listeners suggests that it could pose problems for younger listeners as well. Two studies assessing young children's processing of the natural productions of young children (either their own or another child's) support this intuition.

In an early study, Dodd (1975) examined 2- to 4-year-olds'comprehension of their own mispronounced object labels or those of another child, and compared these with their comprehension of adults' correctly pronounced labels. She found that children were less accurate in choosing the correct object for the children's mispronounced labels (whether their own or another child's) than they were for the adults' correctly pronounced labels. More recently, Cooper, Fecher, and Johnson (2018) used an eye-tracking paradigm to examine the same question in 2.5-year-olds, and found a similar advantage for adult speech. These studies involved many different kinds of errors (from single phoneme errors to multiple errors per word). The presence of multiple errors in a single word results in a significant mismatch between the production and the target representation. Therefore, although such studies indicate that child speech is generally disruptive for child listeners, they do not allow us to evaluate the effects of *specific types* of mispronunciations on children's processing. More precise manipulation of the type of mispronunciation is required to determine whether certain errors are more disruptive than others and whether young listeners' experience or expectations affect their processing of these errors.

Manipulations of single segments have already provided a great deal of insight into toddlers' early lexical processing. For example, whereas words in which a segment is altered by a single phonetic feature are often still mapped to the named object (though significantly less well than correct pronunciations), words with segments altered by two or three features are often not (and may be mapped to a different object). This is true for speech produced by adults (Mani & Plunkett, 2011; White & Morgan, 2008) and by a young school-aged child (Chapter 2, Bernier & White, 2019a). These studies demonstrate that the *degree* of mispronunciation matters. However, as noted above, the mispronunciations produced by children are not random; some mispronunciations are more commonly produced than others (Dodd et al., 2003; Smit, 1993). Therefore, in the present work, we ask whether manipulations of the *type* of mispronunciation (i.e., how commonly they occur) and who is producing them (i.e., a child or an adult) can provide additional insight into how toddlers' experience and expectations affect word processing.

In a recent study, the types of people toddlers had previous experience with influenced how they treated a new individual's speech (Weatherhead & White, 2018). In that study, 16-month-old toddlers heard both familiar (e.g., *dog*) and unfamiliar (e.g., *dag*) pronunciations of known words while viewing images of objects. The words were preceded by photos of either an individual of a familiar race (i.e., their own race) or an individual from a racial group they had little prior experience with. Toddlers' looking behavior differed in these two conditions. In particular, toddlers showed increased looking to unfamiliar objects when they heard unfamiliar pronunciations, such as *dag* (mapping the new labels to the new objects), but only for the familiar-race individual. This suggests that they have an expectation that people who have an unfamiliar appearance (e.g., are from a different racial group) might talk differently. More generally, this shows that toddlers may treat the same speech differently depending on social properties of the speaker.

In the situation above, toddlers were presented with individuals from racial groups they did or did not have prior experience with. In contrast, all hearing children have heard child speech (from either themselves or other children) in addition to adult speech. Thus, it is possible that they have developed expectations about the way children speak, such as the types of mispronunciations they are likely to make. If so, it is possible that mispronunciations commonly made by children will be better tolerated than rarely made ones.

In addition to familiarity with the *type* of changes that are likely to happen, toddlers may also have some expectations about the people who are likely to make them, recognizing that whereas children typically make certain mispronunciations, adults generally do not. Therefore, it is possible that toddlers will be less tolerant of mispronunciations when they are produced by an adult than when they are produced by a child, particularly for those mispronunciations that are common in child speech. As these expectations should be stronger for toddlers with more experience with other children, we additionally compared toddlers with higher and lower amounts of experience with other children.

However, experience-based expectations are not the only factor that could affect toddlers' processing of mispronunciations. For adult speech, asymmetries have been found for infants' and toddlers' processing of certain mispronunciations. For example, some studies (Altvater-Mackensen & Fikkert, 2010; Altvater-Mackensen et al., 2014; Nam & Polka, 2016; van der Feest & Fikkert, 2015) report that toddlers detect single-feature changes in one direction (e.g., /b/→/d/, /v/→/b/), but not in the opposite direction (e.g., /d/→/b/, /b/→/v/). These asymmetries are often (though not always) consistent with differences in input frequency, with changes to frequent sounds (such as /d/) more likely to be detected. Corpus analyses have demonstrated that coronals, such as /d/, are more frequent than other sounds in English. This is true for both speech between adults (Hayden, 1950; Kessler & Treiman, 1997) and speech from adults directed toward young children (Anderson, Morgan, & White, 2003). These asymmetries are also often consistent with underspecification accounts of lexical representation, which predict that listeners are more likely to detect changes to unmarked sounds (e.g., coronals, stops) than the reverse (e.g., labials, fricatives). According to such accounts, only marked information is specified in lexical representations (Lahiri & Reetz, 2010). meaning that mispronunciations of unmarked sounds will not be detected, as the incoming speech does not conflict with the representation.

Based on these types of asymmetries, we might predict a different pattern of results, given that children's mispronunciations are likely to involve changes from less frequent to more frequent sounds, which tend to be produced earlier (Edwards et al., 2015). Therefore, in contrast to our expectation-based predictions, it is possible that commonly made mispronunciations will be detected but that rarely made mispronunciations in the reverse direction will not. However, the majority of studies demonstrating asymmetries have focused on labials (b, v, p, f) and their alveolar counterparts (d, z, t, s). Asymmetries may not be seen for other contrasts (see Tsuji, Fikkert, Yamane, & Mazuka (2016), for evidence of Dutch and Japanese toddlers' sensitivity to a /d/→/g/ change).

In the current study, we assessed 22-month-olds' processing of mispronunciations that were produced by either a child or an adult. Mispronunciations involved a change of a single feature and included three types of change: deaffrication/affrication errors (/t/-/ʧ/), fronting/backing errors involving fricatives (/s/-/ʃ/), and liquid/glide errors (/w/-/ɹ/). We presented toddlers with one of two broad types of mispronunciations – either those that are common in child speech (e.g., /ɹ/→[w]) or those that are infrequent in child speech (e.g., /w/→[ɹ]). All common mispronunciations were listed as being the most frequent error for the youngest age group in Smit (1993). All infrequent mispronunciations did not appear in Smit or were listed as rare.

Toddlers' processing of these mispronunciations was assessed using an intermodal preferential looking task. On each trial, toddlers were shown two objects on a screen (one familiar and one novel). An audio recording (produced by either a first-grade child or a university-aged adult) labeled the familiar object either correctly or with a mispronunciation. Previous work with this type of paradigm has demonstrated that children increase their looking to named objects if they understand the words. The degree to which they increase their looking for mispronounced labels, compared to correctly pronounced labels, provides a measure of their sensitivity to mispronunciations.

If toddlers (a) have expectations about the types of mispronunciations that are common and (b) know that children are more likely to make these mispronunciations, then *common* mispronunciations should be less disruptive, and this should be particularly true for the child speaker. However, given some previously observed asymmetries involving more versus less frequent sounds, it is possible that *infrequent* mispronunciations will not be detected, whereas common mispronunciations will be. If so, this asymmetry should be more evident for the adult speaker. This is because, despite being intelligible to adult listeners, the speech of early school-aged children is more variable than that of adults, particularly for later acquired contrasts such as /s/-/ʃ/ and /w/-/ɹ/ (Klein et al., 2013; Maas & Mailend, 2017; Magloughlin, 2016; Nissen & Fox, 2005).

Experiment 3: Detection of Common and Infrequent Childhood Mispronunciations

Method

Participants. A total of 120 monolingual English-learning 21- to 25-month-olds (M = 22.07 months) were recruited from the Waterloo Region of Ontario, Canada and randomly assigned to one of four conditions (n = 30 per condition). Participants had no more than 10% exposure to another language, and no hearing, vision, or language problems (determined via parent report). Participants were recruited through a database of families who had agreed to be contacted and informed consent was obtained from parents prior to the study. Ten additional toddlers were tested but not included due to fussiness (n = 1), technical issues (n = 2), not knowing at least 7 of the 12 familiar items labeled (n = 2), and not meeting trial inclusion criteria (see details below) on more than half of the correct pronunciation trials (n

= 4) or more than half of all trials (*n* = 1).[5] The study was approved by a research ethics committee at the University of Waterloo.

Design. Each participant was assigned to one Speaker (adult or child) and one Type of Mispronunciation (common or infrequent), resulting in four between-participant experimental conditions (adult common, adult infrequent, child common, and child infrequent). Common mispronunciations were errors most frequently produced by 2- to 3-year-olds (Smit, 1993), and involved a variety of errors types (deaffrication, fronting, liquid gliding). Infrequent mispronunciations involved the same sound pairs as the common mispronunciations but were exchanged in the opposite direction (see Table 6 for examples); these are rarely or not at all produced by 2- to 3-year-olds. Target words were selected based on comprehension norms for 18-month-olds for the MacArthur-Bates Communicative Development Inventories (Fenson et al., 1994; as obtained from the WordBank database [Frank, Braginsky, Yurovsky, & Marchman, 2017]). Due to the limited number of picturable words beginning with /w/, additional items likely to be familiar to toddlers were added in this category. A parental questionnaire after the study was used to verify that the words were familiar to our participants.

There were 24 test trials in total, each with a unique object pair (a list of the familiar objects and their pronunciations is presented in the Appendix). The 20 critical trials presented one familiar object and one novel object. On eight of these trials, the familiar object was labeled correctly; on eight trials, the familiar object was labeled with a small 1-feature

Table 6. Sample pairing of common and infrequent mispronunciations (MP)

	Correct		Mispronounced	
Common MP	raisin	[ɹezɪn]	waisin	[wezɪn]
Infrequent MP	water	[wɑtɚ]	rater	[ɹɑtɚ]

5 The latter 3 criteria exclude filler and /w/-/l/ contrast trials (which were removed due to low item familiarity).

change in the onset consonant; and on four trials the novel object was labeled with a novel word (e.g., *gimble*). The remaining four trials were filler trials with two familiar objects. Novel and filler trials were the same across conditions.

Onsets of the 16 critical familiar-label trials involved four contrasts: /t/-/ʧ/, /s/-/ʃ/, /w/-/ɹ/, and /w/-/l/. For each contrast, toddlers heard two object labels pronounced correctly and two mispronounced. Importantly, correct and mispronounced object labels were counterbalanced across the common and infrequent mispronunciation (MP) conditions, with each child hearing labels beginning with only one member of each contrast. For example, for the common MP condition, toddlers heard only [w] (but not [ɹ]), as in the correct item water and the mispronounced item waisin (for raisin). For the infrequent MP condition, toddlers heard only [ɹ] (but not [w]), as in the correct item *raisin* and the mispronounced item *rauter* (for *water*). All mispronunciations resulted in either a nonword or in a word toddlers are unlikely to know. Onset consonants for the novel words, familiar distractors, and familiar targets in the filler trials, as well as the frames ("Can you find the ___" and "Ooo! A ___"), did not include the critical contrasts.

Stimuli. Stimuli were recorded into Praat (Boersma & Weenink, 2014) in a sound-attenuated room using a Sennheiser e945 microphone connected to a laptop via a blue icicle USB adaptor. In order to have precise control over the mispronunciations, while maintaining the qualities typical of young children's speech, we chose a female first grade student as our child speaker (aged 6;9 [years;months]).To elicit these productions, the child saw an image of the target object on a laptop and repeated the pre-recorded productions of a female adult speaking in a child-directed speech register. The adult stimuli were those of a female graduate student elicited in the same manner. A minimum of three productions per speaker were elicited for each target word. The clearest token (e.g., no unintended mispronunciations

Table 7. Mean acoustic values for child and adult speakers

	Child Speaker	Adult Speaker	Child-Adult Difference
Full Sentence			
Duration (ms)	1722	1728	-5.3
Pitch (Hz)[a]	284	270	14.7
Pitch Variation (SD; Hz)	45.7	69.7	-24.1
Target Word			
Duration (ms)	733	681	51.7
Pitch (Hz)	307	270	37.3
Pitch Variation (SD; Hz)	30.4	70.1	-39.7

a. Pitch values were calculated using a floor of 100 Hz and a ceiling of 600 Hz for both speakers

or extra noise) of each was selected. Stimuli were later adjusted in Praat with all tokens set to the same intensity. They were presented in the testing room through speakers at a comfortable listening volume of 65–70 dB. Table 7 presents the mean acoustic measures of the adult and child productions.

Procedure. Toddlers were tested using the Intermodal Preferential Looking Procedure in a sound attenuated room. The child sat on the parent's lap while the parent listened to music over circumaural headphones. In front of them was a 42-in. widescreen television and two hidden speakers located at the base of the television; both were connected to a computer in an adjacent room. The participants were monitored over closed-circuit video feed that was recorded for later off-line coding; the camera was centrally located beneath the television and hidden behind a black curtain.

There were four blocks of trials, each with one filler trial, one novel label trial, and four critical trials. On these critical trials, the familiar object labels were pronounced correctly on two trials and mispronounced on two trials, with each of the four contrasts (/t/-/ʧ/, /s/-/ʃ/, /w/-/ɹ/, and/w/-/l/) represented. Before each block, the image of a young child or a young adult (depending on speaker condition) was presented for 4 s.

Each test trial consisted of images of two objects, presented on the left and right sides of

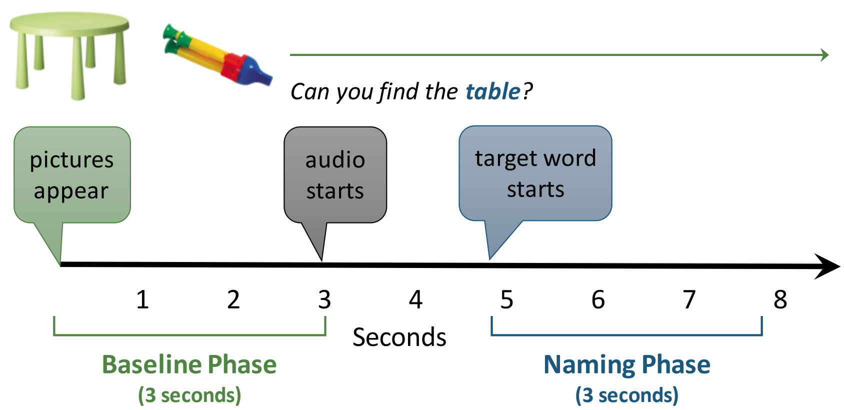


Figure 5. Schematic of a test trial used in Experiment 3.

the screen. The objects were first shown in silence for 3 s; this was used as the baseline phase (see Figure 5). Following the baseline phase, the audio stimulus (e.g., "*Can you find the table?*") began to play. The images remained up for an additional5 s from the start of the audio (for a total trial length of 8 s). We defined a priori a 3-s naming phase that commenced 267 ms (8 frames) after the start of the target word (based on the time needed to program an eye movement and previous convention; e.g., Swingley & Aslin, 2000; White & Morgan, 2008). The dependent measure was the change in the proportion of time toddlers spent looking at each object from the baseline to the naming phase.

Following the testing session, the parent completed a questionnaire on their child's comprehension and production of the experimental items (both familiar and novel), the amount of time their child interacted with other children (siblings, daycare, and playgroups) on a weekly basis, and the number of individuals (toddlers, pre-schoolers, older children, adults) they interacted with throughout the day.

Analysis. Looking behavior was coded off-line using customized software at a rate of 30 frames per second (~33.33 ms/frame). Filler trials, which contained two familiar objects, were excluded from the analysis. For all other trials, which contained one familiar object and one novel object, the proportion of time toddlers spent looking at the *familiar* object (out of the total time spent looking at the two objects) was calculated for both baseline and naming phases. A difference score was then computed (naming minus baseline), which indicates how much toddlers changed their looking to the *familiar* object from baseline to naming. A positive score indicates that they increased their looking to the familiar object following naming, whereas a negative score indicates that they decreased their looking. In this type of paradigm, toddlers significantly increase their looking to a labeled object when they understand the label. The degree to which they show this increase for mispronunciations is typically considered to be a measure of mispronunciation sensitivity (e.g., White & Morgan, 2008). Due to low comprehension of the /w/-/l/ contrast items (e.g., lamp was known by only 51% of participants, with many of the participants recognizing the object as a light), trials pertaining to this contrast were excluded from the analysis.

For a trial to be included in the analysis, the participant needed to look at each of the objects for a minimum of 8 frames (267 ms) during the baseline phase (i.e., a minimum of ~500 ms on-task). This criterion resulted in 10.8% of trials being discarded. Participants also needed to attend to the objects for a minimum of 1 s total during each of the baseline and naming phases. This criterion resulted in 1% of trials being discarded. There were no differences across the four conditions in the overall number of trials excluded (adult common: 11.5%; adult infrequent: 11.3%; child common: 11.9%; child infrequent: 13.3%).

Data were analyzed using linear mixed-effects regression (Baayen et al., 2008) with R's (R Core Team, 2018) *lme4* package (Bates et al., 2015). The between-participant fixed effects of Speaker (adult, child) and MP Type (common, infrequent) were contrast coded; the within-participant fixed effect of Pronunciation (correct, mispronounced) remained treatment coded. Novel labels were assessed separately. The maximum random effects structure that was considered for each analysis included random intercepts for subject and item, as well as by-subject random slopes for Pronunciation, and by-item random slopes for Speaker and Pronunciation. For each analysis, the maximum random effects structure that allowed for convergence was used. The *lmerTest* package (Kuznetsova et al., 2017), which uses Satterthwaite's method for approximating degrees of freedom, was used to assess the significance of model intercepts, factor effects in models with only one fixed effect, and the contrasts of factors with more than two levels (Luke, 2017). Random effects error variance was used as an estimate of σ^2 to calculate Cohen's *d* for model estimates (Brysbaert & Stevens, 2018; Westfall et al., 2014).

Results

We first assessed looking during the baseline phase against chance (0.5). To do this, baseline looking was centered around 0 by subtracting 0.5 from each value, thus making the intercept equivalent to the deviation from chance. A linear mixed-effects regression model was generated with the two between-participant conditions of Speaker (adult, child) and MP Type (common, infrequent), along with their interaction, entered as fixed effects. The random effect structure included random intercepts for subject and item (i.e., the specific object being labeled) and by-item random slopes for Speaker. Because the pronunciation of the forthcoming label (correct or mispronounced) should not have affected baseline performance, a random slope for pronunciation was not included. Overall, toddlers spent equivalent amounts of time during baseline on the familiar and novel objects (β_0 = 0.011, β_0 *SE* = 0.017), $t(15.8) = 0.665$, $p = .515$, $d = 0.06$. Comparison of maximum likelihood models showed that this did not differ across the four between-participant experimental conditions (adult common, adult infrequent, child common, and child infrequent), $\chi^2 \leq 1.01$, $p \geq .315$.

Our main analyses involved toddlers' change in looking (difference scores) for correct (CP) and mispronounced (MP) labels of familiar objects. As a reference point, we first assessed performance on correctly pronounced labels across the four between-participant experimental conditions. Next, we examined whether the MP Type (common or infrequent) and Speaker (adult or child) influenced toddlers' processing of the mispronounced labels. As each type of mispronunciation involved three different kinds of contrasts (/t/-/ʧ/, /s/-/ʃ/, and /w/-/ɹ/), we subsequently looked at performance on each of the contrast pairs independently to determine whether the pattern we observed overall held for each contrast. Finally, we explored whether the patterns differed for toddlers with differing amounts of exposure to other children.

***Correctly pronounced labels*.** To compare toddlers' recognition of correctly pronounced (CP) labels in the four experimental conditions, a model was generated for CP trials with Speaker (adult, child), MP Type (common, infrequent), and their interaction entered as fixed effects, along with random intercepts for subject and item. This model showed that toddlers significantly increased their looking to the familiar object from baseline (β_0 = 0.156, β_0 *SE* = .019), $t(12.3) = 8.253$, $p < .001$, d = 0.5. Comparison of maximum likelihood models showed that this did not differ across conditions, χ^2s ≤ 2.25, *p*s ≥ .134.

***Type of mispronunciation and speaker*.** Next, we assessed the effect of mispronunciations with a model that included Pronunciation (CP, MP), Speaker (adult, child), MP Type (common, infrequent), and their interactions as fixed effects. The maximum random effects structure included random intercepts for subject and item, as well as by-subject random slopes for Pronunciation and by-item random slopes for Speaker. Comparison of maximum likelihood models revealed a significant effect of Pronunciation, $\chi^2(1) = 26.13$, $p < .001$, that did not interact with Speaker, $\chi^2(1) = 0.42$, $p = .519$. Thus, consistent with prior work (Chapter 2, Bernier & White, 2019a; White & Morgan, 2008), model estimates showed an overall mispronunciation penalty ($\beta = -0.098$, β *SE* = 0.018, $d = -0.31$) that was similar in size across child and adult speakers (β = −0.023, β *SE* = 0.036, *d* = −0.07). Importantly however, there was a 3-way interaction between Pronunciation, Speaker, and MP Type, $\chi^2(1) = 3.92$, $p = .048$ (β = −0.161, β *SE* = 0.082, *d* = −0.50). No other effects reached significance, χ^2s ≤ 2.53, *ps* ≥ .112.

To examine the 3-way interaction (see Figure 6), models were generated separately for each Speaker, with Pronunciation, MP Type, and their interaction entered as fixed effects. The maximum random effects structure included random intercepts for subject and item for the models of both speakers, as well as by-subject and by-item random slopes for Pronunciation for the child speaker model.

Adult speaker. Model comparisons for the Adult Speaker conditions revealed that, in addition to an expected overall MP penalty, $\chi^2(1) = 12.79$, $p < .001$ (β = −0.087, β *SE* = 0.024, *d* = −0.28), the size of the penalty significantly differed across MP Types, $\chi^2(1) = 5.79$, $p = .016$. The mispronunciation penalty was substantially smaller for infrequent MPs than common MPs (β = 0.143, β *SE* = 0.057, *d* = 0.46).

To further examine this difference, models were generated for each MP Type with Pronunciation as a fixed effect, along with random intercepts for subject and item; by-subject random slopes for Pronunciation were included for the infrequent MP condition only. Intercept-only models (i.e., no fixed effects, with only random intercepts for subject and item)

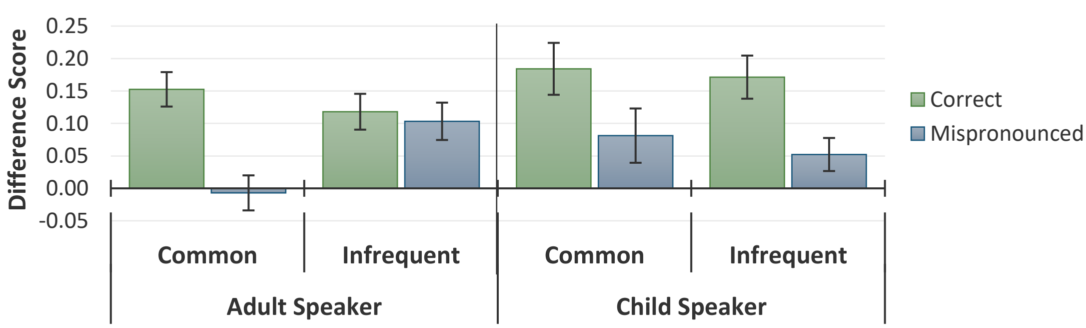


Figure 6. Model estimates of the difference scores for familiar object labels for the four experimental conditions in Experiment 3. Error bars represent standard error.

were used to assess toddlers' change in looking from baseline for the mispronounced labels. For common mispronunciations, toddlers did not change their looking from baseline (β = −0.007, β *SE* = 0.027), $t(5.0) = -0.244$, $p = .817$, $d = -0.02$, which was significantly different from what they did for correctly pronounced labels (β = −0.159, β *SE* = 0.037), $t(296.2) = -4.337$, $p < .001$, $d = -0.48$. Thus, toddlers showed a large mispronunciation effect when an adult speaker produced words with common childhood mispronunciations. For infrequently mispronounced labels, however, toddlers significantly increased their looking to the familiar object from baseline (β = 0.103, β *SE* = 0.029), $t(6.0) = 3.588$, $p = .012$, $d = 0.36$, doing so as much as they did for correctly pronounced labels (β = −0.015, β *SE* = 0.036), $t(8.0) = -0.403$, $p = .697$, $d = -0.05$. In other words, infrequent mispronunciations produced by an adult were treated similarly to correct pronunciations.

Child speaker. Model comparisons for the Child Speaker conditions revealed only an overall effect of Pronunciation, $\chi^2(1) = 8.01$, $p = .005$ (β = −0.110, β *SE* = 0.036, $d = -0.32$), and no interaction with MP Type, $\chi^2(1) = 0.06$, $p = .813$ (β = −0.015, β *SE* = 0.072, $d = -0.05$). Models for each type of mispronunciation showed that toddlers exhibited similarly sized MP penalties for common MPs (β = −0.102, β *SE* = 0.0552, $d = -0.30$) and infrequent MPs (β = −0.119, β *SE* = 0.047, $d = -0.38$). An intercept-only model for MP labels (taken across both types of MPs) showed that toddlers significantly increased their looking to the familiar objects when hearing these mispronunciations (β = 0.067, β *SE* = 0.023), $t(11.6) = 2.903$, $p = .014$, $d = 0.21$.

Based on this difference across speakers, we directly compared toddlers' processing of mispronunciations for the Adult and Child Speaker conditions. Models for each MP Type included Pronunciation and Speaker, and their interaction as fixed effects, along with random intercepts for subject and item; by-subject random slopes for Pronunciation were also included in the common MP model. Model comparisons showed that the interaction effect for common MPs was not significant, $\chi^2(1) = 1.01$, $p = .295$ ($\beta = 0.058$, β *SE* = 0.056, $d = 0.17$), whereas the interaction effect for infrequent MPs was, $\chi^2(1) = 4.93$, $p = .026$ ($\beta = -0.104$, β *SE* = 0.047, $d = -0.35$). In other words, although the mispronunciation penalties differed across speakers (common: $d = -0.48$ vs. -0.30 for adult and child speakers respectively; infrequent: $d = -0.05$ vs. -0.38, respectively), only the difference for the infrequent mispronunciations was significant, and this was driven by a reduced penalty for the adult speaker. We will address this result further in the Discussion section.

Individual contrasts. We next considered whether these patterns were similar across the three contrasts assessed (/t/-/ʧ/, /s/-/ʃ/, and /w/-/ɹ/). To examine this, we first generated models separately for each contrast for the four experimental conditions, with Pronunciation

Table 8. Model estimates (SE) of the difference scores for each contrast as a function of Speaker, along with effect sizes (Cohen's d) of the mispronunciation (MP) effects.

	Common Mispronunciations			Infrequent Mispronunciations		
	CP	**MP effect[a]**	***d***	**CP**	**MP effect[a]**	***d***
Adult Speaker						
/t/-/ʧ/	0.201 (0.045)	-0.190 (0.064)	-0.58	0.186 (0.069)	-0.082 (0.097)	-0.32
/s/-/ʃ/	0.101 (0.061)	-0.126 (0.088)	-0.38	0.106 (0.040)	-0.010 (0.056)	-0.03
/w/-/ɹ/	0.156 (0.048)	-0.161 (0.064)	-0.48	0.069 (0.043)	0.037 (0.056)	0.12
Child Speaker						
/t/-/ʧ/	0.207 (0.102)	-0.143 (0.143)	-0.42	0.220 (0.051)	-0.150 (0.069)	-0.49
/s/-/ʃ/	0.178 (0.046)	-0.066 (0.058)	-0.22	0.095 (0.054)	-0.087 (0.076)	-0.28
/w/-/ɹ/	0.153 (0.068)	-0.088 (0.094)	-0.28	0.073 (0.073)	-0.118 (0.102)	-0.37

a. the MP effect indicates the difference in looking behavior for the mispronounced (MP) labels relative to the correct (CP) labels; it represents both the size and direction of the difference entered as a fixed effect, along with random intercepts for subject and item (see Table 8 for model estimates). Regardless of Speaker or MP Type, toddlers exhibited larger mispronunciation effects for the /t/-/ʧ/ contrast than they did for the /s/-/ʃ/ and /w/-/ɹ/ contrasts. Even toddlers who heard infrequent mispronunciations from the adult speaker detected the /t/-/ʧ/ mispronunciations. To assess the significance of this observed difference, a model was generated over all participants with Pronunciation and Contrast Pair entered as fixed effects (Contrast Pair was coded to examine, first, /s/-/ʃ/ vs. /w/-/ɹ/, and, second, these two contrasts vs. /t/-/ʧ/). The maximum random effects structure included random intercepts for subject and item, as well as by-subject random slopes for Pronunciation and by-item random slopes for Speaker. This analysis showed that the sizes of the MP effects for /s/-/ʃ/ and /w/-/ɹ/ were not different from one another ($\beta = -0.011$, β $SE = 0.042$), $t(1061) = -0.262$, $p = .794$, $d = -0.03$, whereas the size of the MP effect for /t/-/ʧ/ was marginally larger than the MP effect of the other two ($\beta = 0.064$, β $SE = 0.036$), $t(1061) = 1.801$, $p = .072$, $d = 0.20$. Together, this indicates that mispronunciations involving /t/ and /ʧ/ led to the greatest decreases in looking, irrespective of direction and who produced them.

Experience with other children. To explore the effects of experience with other children, we compared the performance of the 50 toddlers who had the least experience with other children in group activities to that of the 50 toddlers who had the most experience (≤ 10 hr/week [$M = 3.3$; 0.8 hr/week in daycare] vs. ≥ 35 hr/ week [$M = 41.4$; 39.7 hr/week in daycare], equally represented from each condition). The 20 participants who made up the middle range of experience were excluded in order to get a clean separation in experience between the groups. Because splitting the data this way resulted in a small number of participants ($n = 12$ or 13) per Experience Level for each of the four conditions, we report only effect sizes.

We first generated models separately for High and Low Experience groups for the four experimental conditions, with Pronunciation entered as a fixed effect, along with random intercepts for subject and item (see Table 9 for model estimates). For the adult speaker, toddlers showed a greater penalty for common than infrequent mispronunciations, regardless of their experience with other children. For mispronunciations produced by the child speaker, there were differences as a function of experience group, but only for the common mispronunciations. Toddlers with more child experience who heard common mispronunciations from the child speaker showed a smaller mispronunciation penalty than the other three groups who heard the child speaker (high experience/infrequent, low experience/common and infrequent); the latter three groups had equivalently sized mispronunciation penalties.

Novel labels. Finally, to examine novel label processing, a model was generated for novel label trials with the two between-participant factors of Speaker (adult, child) and MP Type

Table 9. Model estimates (SE) of the difference scores for each Speaker as a function of Experience, along with effect sizes (Cohen's d) of the mispronunciation (MP) and novel label effects.

	Common Mispronunciations			**Infrequent Mispronunciations**			**Across MP Type**	
	CP	**MP effect[a]**	***d***	**CP**	**MP effect[a]**	***d***	**Novel Label**	***d***
High Experience								
Adult Speaker	0.153 (0.055)	-0.148 (0.078)	-0.42	0.107 (0.038)	-0.003 (0.052)	-0.01	-0.070 (0.063)	-0.20
Child Speaker	0.101 (0.054)	-0.034 (0.071)	-0.10	0.166 (0.042)	-0.099 (0.057)	-0.32	-0.058 (0.064)	-0.16
Low Experience								
Adult Speaker	0.202 (0.035)	-0.165 (0.051)	-0.57	0.142 (0.044)	-0.041 (0.051)	-0.14	-0.051 (0.055)	-0.17
Child Speaker	0.208 (0.050)	-0.101 (0.061)	-0.36	0.146 (0.058)	-0.110 (0.082)	-0.33	0.022 (0.056)	0.07

a. the MP effect indicates the difference in looking behavior for the mispronounced (MP) labels relative to the correct (CP) labels; it represents both the size and direction of the difference (common, infrequent) and their interaction entered as fixed effects, along with random intercepts for subject and item. This analysis showed that, overall, toddlers did not map novel labels to novel objects ($\beta_0 = -0.042$, β_0 *SE* = 0.051), $t(3.2) = -0.828$, $p = .465$, $d = -0.12$. Maximum likelihood model comparisons showed no difference across the four experimental conditions, χ^2s ≤ 0.720, *p*s $\geq .396$.

However, prior work has shown that toddlers with different amounts of experience with other children may process novel labels differently (Chapter 2, Bernier & White, 2019a). Therefore, we also report novel label performance as a function of experience. The rightmost columns of Table 9 show that toddlers with more child experience looked toward the novel object upon hearing a novel label regardless of which speaker they heard produce it. However, toddlers with limited experience looked toward the novel object only when the label was produced by an adult speaker, and not when it was produced by a child speaker. This differential responding is consistent with the possibility that experience with other children affects toddlers' novel label processing.

Discussion

We assessed toddlers' processing of mispronunciations as a function of the type of mispronunciations they heard (commonly or infrequently occurring in children's speech) and the speaker who produced them (a child or an adult). Overall, toddlers showed similarly sized penalties for mispronunciations from child and adult speakers when the type of mispronunciation was not taken into account. However, when the type of mispronunciation was considered, toddlers' processing of common and infrequent mispronunciations differed for child and adult speakers, but not as predicted by an expectation-based account.

***Type of mispronunciation: Adult speaker*.** On an expectation-based account, there should be no difference in the processing of common versus infrequent child mispronunciations if they are produced by adults. However, our results show that these two types of mispronunciations are treated differently, with those that are common in children's speech resulting in a large mispronunciation penalty and those that are rare being treated equivalently to correctly pronounced labels. This pattern is consistent with some previously observed asymmetries that have been attributed to differences in input frequency, lexical underspecification, or the perceptual stability of stops (Altvater-Mackensen & Fikkert, 2010; Altvater-Mackensen et al., 2014; Nam & Polka, 2016; van der Feest & Fikkert, 2015). For example, some have argued that children should be less likely to detect a change from a more frequently heard sound to a less frequent one (such as '*soap'* mispronounced as '*shoap'*) because the latter may be less well represented and thus more difficult to identify. In contrast, children should be more likely to detect a change to a more frequent sound (such as '*shoe'* mispronounced as '*sue'*) because the onset of the mispronunciation should be more robustly represented, and thus easily recognized as incorrect. A similar asymmetry is predicted on underspecification accounts, where changes from unmarked – typically more frequent – sounds should be more difficult to detect.

An alternate explanation for the asymmetries we observed lies with perceptual confusability. To assess this possibility, we looked at how often the onsets of the mispronounced (MP) labels (i.e., the sounds toddlers actually heard) are confused with the correct (CP) onsets. The confusability tables of Wang and Bilger (1973), which include all of the sounds we assessed, demonstrate that, for both the /t/-/ʧ/ and /s/-/ʃ/ contrasts, the infrequent MP onsets (/ʧ/ and /ʃ/) are more likely to be misperceived as the corresponding CP onsets (/t/ and /s/) than the reverse. For example, the [ʧ] in *chable* (for table) is more likely to be perceived as the correct onset /t/ than the [t] in *tair* (for chair) is to be perceived as the correct onset /ʧ/. Therefore, toddlers' similar treatment of infrequent mispronunciations and correct pronunciations may have been in part due to confusability (although the asymmetry for /w/-/ɹ/ in our data is not mirrored by an asymmetry in confusability in the Wang and Bilger dataset). However, these confusion matrices are based on adult judgments, so it is important to also consider confusability from a child listener's perspective.

Nishi, Lewis, Hoover, Choi, and Stelmachowicz (2010) examined consonant confusability for 4- to 5- year-old and adult listeners. They found that consonants were, overall, more confusable for the children than they were for adults, and that children exhibited some of the same asymmetries in confusability that adults do (e.g., [ʃ] was more likely to be misperceived as /s/ than the reverse). This provides additional support that, on at least some of the trials, toddlers who heard infrequent mispronunciations by the adult speaker may have misperceived the onsets as the correct sounds. This may in part explain why toddlers in this condition did not show a mispronunciation penalty.

It is also possible that toddlers' own productive abilities contributed to the asymmetry in performance. Indeed, infants' listening preference for speech sounds has been shown to be influenced by their own production abilities (DePaolis, Vihman, & Keren-Portnoy, 2011; Majorano, Vihman, & DePaolis, 2014). Sounds that toddlers can produce themselves are likely to be particularly robustly represented and processed more easily in the input. This should make the mismatch between the input and their lexical representation easier to detect. By design our common mispronunciations are sounds that toddlers should be able to produce well given their early acquisition (Dodd et al., 2003).

One factor that has been argued to affect toddlers' phonological development is type frequency (Edwards et al., 2015), with sounds that occur across more words in the child's productive lexicon thought to be more robustly represented. We therefore examined the number of different words (type frequency) in the MacArthur-Bates Communicative Development Inventories: Words & Sentences (Fenson et al., 1994) that begin with each of our target sounds (excluding onset clusters). We found that there are approximately 2.3 times as many words beginning with our infrequently mispronounced onsets (/t/, /s/, /w/) as with their commonly mispronounced counterparts (/ʧ/, /ʃ/, /ɹ/) and that the pattern is similar for each of our contrast pairs (/t/-/ʧ/, /s/-/ʃ/, /w/-/ɹ/). Importantly, a summary of the data for WordBank's American English-learning 20- to 25-month-olds (n = 1,855) shows that toddlers this age are producing words with both types of onsets at about the same rate (Frank et al., 2017). This means that children are likely producing the sounds /t/, /s/, and /w/ in a greater number of phonological contexts (i.e., across more different words), which may contribute to more robust knowledge of these sounds. Further work is needed to better assess this possibility, as it requires an assessment of individual participants' productive abilities.

Finally, we note that a similar failure to detect infrequent mispronunciations was also found recently with older children. Krueger, Storkel, and Minai (2018) presented 5-year-olds with images of two objects – a familiar (real) object and a novel (non) object – along with a recording of an adult labeling the real object either correctly, with a common substitution, or with an uncommon substitution (e.g., thumb pronounced correctly as *fumb* [a common error] or as *shum* [an uncommon error]). In other words, in contrast to our manipulation of frequency (which used the same onset pairs, changing only the direction in which the exchange occurred), Krueger et al. used a particular set of objects whose labels were changed in different ways for common and uncommon mispronunciations. In their eye-tracking experiment, after children heard an object label mispronounced with a common substitution (e.g., *fumb* for thumb), they looked equally at both objects. However, after hearing the object label produced with an uncommon substitution (e.g., *shum* for thumb), children looked more to the real object. So, like our 22-month-olds, children in Krueger et al. showed chance looking behavior when they heard an adult produce a familiar label with a common substitution but looked more to the familiar object when they heard an adult produce the label with an uncommon substitution. However, the object labels in their study involved a variety of different contrasts, the majority of which have not specifically been shown to have asymmetric effects in perception. Therefore, further work is needed to determine whether the same factors are driving the asymmetry in the two studies.

Type of mispronunciation: Child speaker. Turning now to the child speaker, an expectation-based account predicts that toddlers should recognize that only children typically mispronounce words, and that only certain kinds of mispronunciations are likely to occur. In other words, toddlers should be more tolerant of common mispronunciations from a child speaker. However, given previously observed asymmetries when toddlers listen to adult speech, we might alternatively predict that they should show the same (though possibly less clear) types of asymmetries when listening to child speech. In this case, toddlers should detect common mispronunciations but not infrequent ones. Our results did not align with either of these possible outcomes. Instead, we found that, toddlers' treatment of common and infrequent mispronunciations was equivalent when they were produced by a child. The primary difference across speakers, in fact, was that toddlers showed a reduced penalty for infrequent mispronunciations produced by an adult speaker.

If expectations are not driving toddlers' responses, why don't they show the same asymmetry for the child speaker as they do for the adult speaker? One potential reason is the nature of the child speaker's productions. Even at the age of 6 years, when children produce few overt errors in their speech (Dodd et al., 2003), their productions continue to be less accurate, showing more variability than those of adults. Compared to adults, early school-aged children have been shown to have a more variable /s/ (Koenig et al., 2008; Munson, 2004), a less distinguishable /s/ and /ʃ/ (Maas & Mailend, 2017; Nissen & Fox, 2005), less accurate /ɹ/ (Klein et al., 2013; Magloughlin, 2016) and /l/ (Lin et al., 2016), and to have a larger vowel space (Hillenbrand et al., 1995; Lee et al., 1999), among other differences. In terms of our stimuli, this means that the child speech should be less accurate, especially for later acquired sounds such as /ʃ/ and /ɹ/ (likely making them more similar to /s/ and /w/, respectively). However, this would predict an overall reduction in the size of the mispronunciation penalties for the child speaker (compared to the adult speaker). Instead, we found that, although toddlers were somewhat worse at detecting common mispronunciations (compared to the adult speaker), they were actually *better* at detecting infrequent ones. Additional work is needed to better understand this difference.

Finally, although the type of mispronunciation did not matter with the child speaker, it is important to note that all of our mispronunciations involved sounds that are frequently exchanged with one another by children (albeit typically in only one direction). It is therefore possible that toddlers treat mispronunciations involving sounds that are frequently exchanged differently than mispronunciations involving sounds that are rarely exchanged. Additional work would need to be done to determine whether toddlers' treatment of mispronunciations is different for contrasts such as /f/-/h/ (which are rarely or never exchanged with one another).

Individual contrasts. Toddlers did not respond equally for all contrasts. Instead, they showed larger penalties for mispronunciations involving /t/-/ʧ/ compared to those involving /s/-/ʃ/ or /w/-/ɹ/ (which were treated similarly). Even in the adult infrequent condition, where there was no mispronunciation penalty overall, toddlers showed a penalty for /t/-/ʧ/ mispronunciations. This suggests that the /t/-/ʧ/ contrast is processed differently than the other contrasts.

We begin by considering whether perceptual confusability can explain this pattern. In the confusion matrices of Wang and Bilger (1973), only the /t/-/ʧ/ pair showed a strong directional asymmetry in the probability of a misperception (i.e., 1.7% vs. 8.7% of the time); misperceptions for the other contrasts occurred equally often in both directions (i.e., 3.3%–4.9% of the time). Therefore, perceptual confusability does not explain why we found directional differences for the /s/-/ʃ/ and /w/-/ɹ/ contrasts, nor does it explain the stronger overall effect for the /t/-/ʧ/ contrast.

Another possible explanation for the different pattern across contrasts is that /t/-/ʧ/ errors might be less frequently encountered in the input. Because mispronunciations in adults are relatively rare, we consider the frequency with which 2-year-olds produce these errors. By design, our common mispronunciations occur more frequently than our infrequent ones, which are extremely rare. Smit (1993) found that infrequent /t/→[ʧ] errors never occurred and infrequent /s/→[ʃ] and /w/→[ɹ] errors occurred less than 1% of the time. For common mispronunciations, Smit found that /ʧ/→[t] and /ʃ/→[s] occur at roughly the same rate, whereas /ɹ/→[w] errors occur much more often. Thus, frequency of occurrence also cannot explain the stronger effects of /t/-/ʧ/ mispronunciations in *both* directions, nor can it explain why toddlers showed the same directional differences for /s/-/ʃ/ and /w/-/ɹ/ mispronunciations.

A more likely possibility is that the stronger effect for the /t/-/ʧ/ pair is due to the fact that /t/ is an early acquired sound (Dodd et al., 2003), and is therefore likely to be very robustly represented. Indeed, stops in general are particularly well represented, and may be more stable in children's production and perception (Altvater-Mackensen et al., 2014).

Additionally, for /t/, Nishi and colleagues (2010) found that 4- to 5-year-olds had adultlike perception and low error rates. Therefore, it may be particularly salient to toddlers when a sound is changed to (or from) /t/. This is consistent with the high sensitivity to these common mispronunciations in our data (e.g., the [t] in *tair* [for chair] was unlikely to have been misperceived as the correct sound /ʧ/). One final consideration is that /ʧ/ is a more complex sound than the others, one that is characterized by a rapid sequence of /t/ and /ʃ/ with acoustic cues of both stops and fricatives. It is possible that the addition or deletion of frication for the /t/-/ʧ/ contrast is easier to detect than the complete substitutions of the other mispronunciations.

Experience with other children. Although we did not find patterns that were consistent with an expectation-based account when we considered our toddlers as a whole, because toddlers' expectations about children's mispronunciations should be influenced by experience, we also compared toddlers with the most experience with other children in group settings to those with the least experience. Regardless of their experience, toddlers showed the same pattern when mispronunciations were produced by the adult speaker (with a larger penalty for common mispronunciations). For mispronunciations produced by the child speaker, however, toddlers with high and low levels of experience trended toward treating common mispronunciations differently – toddlers with more experience showed a smaller penalty (possibly reflecting greater tolerance for the mispronunciations). Similarly, for the processing of novel labels, we found no effect of experience when the labels were produced by an adult. However, when the labels were produced by a child, only toddlers with more child experience looked toward the novel object; those with less experience did not change their looking from baseline. Together this suggests that experience with other children in group settings has some influence on toddlers' processing of child speech and may additionally influence their willingness to learn a novel label from a child.

However, given the small sample size for each experience group, it remains possible that these results are due to Type I error and that toddlers varying in their experience with other children do not differ in their processing of child speech. If so, this would suggest that, at this age, even extensive experience with other children is not enough to build expectations about how children speak (compared to adults). Although young toddlers have been shown to have a general expectation that people who look different may speak differently (Weatherhead & White, 2018), expectations about speakers of different age groups may not emerge until later in development. By the time children are 5 years old, they appear to have developed some familiarity with the types of mispronunciations that occur in child speech, demonstrating greater tolerance for common childhood mispronunciations than infrequent ones in selection tasks (Krueger et al., 2018). However, that work used adult speech, so it did not directly compare children's treatment of mispronunciations from speakers of different ages. Further work is needed to determine if and when children develop differential expectations for child versus adult speakers, and how this is affected by experience.

Conclusion. This study is the first to directly compare toddlers' processing of mispronunciations produced by adults and children. The results both reinforce and add to previous investigations of toddlers' mispronunciation processing. We found that toddlers treated common and infrequent childhood mispronunciations differently when they were produced by an adult, with infrequent mispronunciations having a minimal effect on processing. This pattern is consistent with some previously reported perceptual asymmetries. In contrast, toddlers had a similar response to these two types of mispronunciations when they were produced by a child. This difference between child and adult speech will be important to explore in the future, as it may provide additional insight into the factors contributing to perceptual asymmetries.

We additionally found that errors involving /t/-/ʧ/ produced a larger mispronunciation penalty than did errors involving /s/-/ʃ/ and /w/-/ɹ/, a result that is not unexpected given that /t/ is likely better known than the other sounds. More generally, toddlers seemed to be better at detecting mispronunciations in adult speech that involved changes to sounds that they are likely to be able to produce themselves. Finally, the number of hours spent with other children appeared to have some influence on toddlers' performance. Those with more experience appeared more tolerant of commonly produced childhood mispronunciations, and more willing to map novel labels to novel objects for a child speaker. Our results show a complex interplay of speaker, type of mispronunciation, specific contrast, and amount of experience with other children in driving toddlers' processing of mispronunciations.

Chapter Four: Expectations about a Common Childhood Mispronunciation

As adults, we have developed expectations about the way different people speak. Many of these expectations are driven by accumulated experiences with the speech patterns of different individuals, groups, and even contexts. In a particular situation, we may apply these expectations because of bottom-up cues (e.g., detecting the cues corresponding to a particular accent) or because of top-down information (e.g., knowledge that a person belongs to a particular group). For example, when an adult is primed with a nationality (e.g., Canadian vs. US Detroiter; Australian vs. New Zealander), they will process a person's speech in ways that are consistent with the primed nationality (Hay et al., 2006; Niedzielski, 1999). Often, these expectations help us process speech in an efficient manner. But how and when these expectations develop is unknown. Though some work has begun asking this question in regards to foreign and accented speech (Uttley et al., 2013; Weatherhead & White, 2018), in the present study, we ask whether 2- and 3-year old children have expectations about the speech of adults vs. children (social groups that differ in their speech patterns), and about the role of experience in building these expectations.

There are indications that even infants have some very general expectations about the ways different types of people speak. For example, 6-month-old infants match unfamiliar languages to unfamiliar race speakers (Uttley et al., 2013). And recent work has shown that 16-month-olds have expectations about the way people of different racial backgrounds might pronounce words (Weatherhead & White, 2018). In this study, infants heard both natively accented and unfamiliarly accented words from a speaker of their own race or a different race. When the infants saw the *familiar*-looking speaker (own race), they initially treated the new pronunciations as novel words. In contrast, when they saw the *unfamiliar*-looking speaker (different race), they were initially unsure of how to interpret either type of pronunciation. This study indicates that infants have some expectations about how adults might sound based on whether they are like the people around them or different from the people they typically encounter. In particular, they appear to expect familiar-looking individuals to talk in familiar ways, but to not have the same expectation for unfamiliar-looking individuals.

Another type of speaker that would be particularly helpful for young children to have expectations about would be other children. Early on, children's productions are characterized by a number of deviations from adult targets, including (among others) substitutions of one sound for another (e.g., *fumb* for thumb) and omissions of sounds (or syllables) altogether (e.g., *nake* for snake). These kinds of deviations can lead to low intelligibility for naïve adult listeners (Flipsen, 2006; Hodson & Paden, 1981). By 4 years the majority of these larger deviations have decreased, though some sounds, such as /ɹ/ and /θ/, remain challenging for many children to produce accurately until 6 years or later (Dodd et al., 2003).

To our knowledge, only two studies other than those in this dissertation have thus far directly compared very young children's processing of the speech of other children with their processing of adult speech. In a first study, Dodd (1975) tested 2- to 4-year-olds' comprehension of their own productions of object labels, as well as those of another child, and an adult. She found that children were less accurate in choosing the named object when hearing labels produced by a child (including their own productions) compared to labels produced by an adult. More recently, Cooper, Fecher, and Johnson (2018) used an eye-tracking paradigm to test the same question with 2½-year-olds, and similarly found an advantage for adult speech. However, this disadvantage for processing child speech may be gone by the time child speakers have reached school age (Chapter 2; Bernier & White, 2019a).

These studies indicate that the speech of very young children, with its many errors, deviates too much from adult speech for toddlers to recognize the intended words. In fact, even naïve adult listeners have difficulty processing the speech of very young children (Flipsen, 2006; Hodson & Paden, 1981) unless they have experience listening to children speak (Munson et al., 2012). Because these studies relied on labels containing many errors, their deviations are too extreme for us to ask whether young children have *specific expectations* about the way other children speak. One such expectation would be the recognition that children commonly mispronounce /ɹ/ as [w]. To ask this question, children would need to be presented with labels that contain only a single speech error (e.g., rainbow pronounced as *wainbow*). If children have an expectation about this mispronunciation, this should help them infer that an unfamiliar child's production of *wainbow* refers to the object rainbow. In contrast, specific expectations based on previous experience should not help young children infer the intended referent in cases where the speaker is producing a mispronunciation that is rare in child speech.

Following this line of reasoning, Krueger, Storkel, and Minai (2018) assessed 5-year-old children's processing of object labels produced by an adult or another child. Labels produced by the adult speaker were correctly pronounced or mispronounced with either a common or a rare childhood mispronunciation (e.g., leaf pronounced as *weaf* [an error that is common in child speech] or *yeaf* [an error children rarely make]). The labels produced by other children were only pronounced correctly or with a natural (i.e., common) mispronunciation. In a selection task, Krueger et al. (2018) found that 5-year-olds detected common mispronunciations from both the adult and child speakers, thus demonstrating sensitivity to phonetic detail for both types of speakers. But they also found that the 5-year-olds showed a smaller penalty when the adult produced common mispronunciations than when the adult produced rare ones. For example, children were more likely to select the image of a leaf after hearing the adult use the label *weaf* than after hearing the adult use the label *yeaf*. That children distinguished between common and rare mispronunciations (at least when they were produced by an adult speaker) suggests that they were influenced by their previous experiences with child speech (either their own or other children). However, there was no direct comparison of the treatment of common and rare mispronunciations produced by a child speaker.

In work that directly compared common and rare mispronunciations from a child speaker, 21- to 25-month-olds treated common and rare mispronunciations (e.g., /ɹ/→[w] vs /w/→[ɹ]) similarly (Chapter 3; Bernier & White, 2019b). This result suggests that young children may not have specific expectations about the errors that other children will make. However, children's perception of speech errors in familiar words may be influenced by the manner in which they have heard the particular words before. Testing children on their perception of newly learned words is an alternative approach that has the potential to more clearly reveal the use of expectations, as it is independent of experience with specific words. If young children expect other children to make certain kinds of mispronunciations, then they should be able to "undo" a child's common speech error, even for a newly learned word.

In the current study, we asked whether 2- and 3-year-old children would undo a child-typical /ɹ/→[w] mispronunciation for a novel word when it was produced by a 5-year-old who had produced familiar words with this same error. To do so, we trained young children on a novel word (*roogie*) from an adult, and subsequently tested them on a mispronunciation of that word (*woogie*), from the same adult and a child. Prior to testing, the participants were familiarized to the adult and the child speakers producing familiar words starting with /ɹ/.

The adult labelled the familiar objects correctly, whereas the child labelled the same familiar objects with an /ɹ/→[w] mispronunciation.

Experiment 4: Expectations in 2-year-olds

As a first step to asking whether toddlers have expectations about child speech, we ask whether 2-year-olds will be able to undo a common childhood mispronunciation for a newly learned word. If toddlers recognize the child's /ɹ/→[w] mispronunciation at test and undo it, then they should map the mispronounced label *woogie* to the trained object for a child speaker. In contrast, when the adult uses the same label, they should infer (via a process of disambiguation) that she is referring to the untrained (never-before-seen) object, because she used a different word (*roogie*) to label the trained object and because she has demonstrated that she correctly pronounces /ɹ/. Alternately, if 2-year-olds fail to undo the child's /ɹ/→[w] mispronunciation, then they should interpret the child's *woogie* as a label for the untrained object for the child speaker as well.

Because experience should be key to the development of these kinds of specific expectations, we also considered the role of participants' experience with other child speakers. We hypothesized that young children with more experience hearing the speech of other children would be more likely to have developed these expectations and consequently be more likely to undo the child's mispronunciation and map it to the trained object.

Method

Participants. A total of 32 monolingual English-learning 2-year-olds (mean age 27 months; range 25.4-27.9) were recruited from the Waterloo Region of Ontario, Canada. Four additional 2-year-olds were tested but not included due to not meeting trial inclusion criteria (see below) on both test trials.

Stimuli. To ensure the child speaker (aged 5 years 6 months) was as comfortable as possible with the recording setup, videos for familiarization and training stimuli were recorded using an iPad mini 2 placed on a level surface with an attached Sony headset with built in microphone. Both the adult and child speakers were recorded in a central position looking towards the camera while sitting behind a light neutral-colored table, in front of a light neutral-colored wall (see Figure 7 for images of the child speaker).

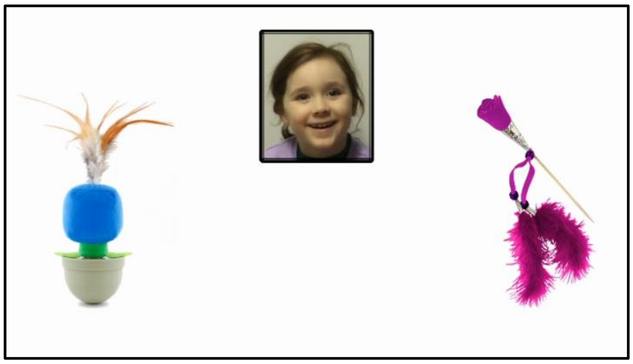
Familiarization: Both speakers were recorded labelling the same 4 familiar objects (rooster, rabbit, rattle, rainbow) in a variety of carrier phrases. The adult speaker labelled the objects correctly; the child speaker labelled the objects with an /ɹ/→[w] mispronunciation (see Appendix Table C1 for the scripts used in the study). Training: The adult speaker was recorded labelling the novel training object *roogie* in a variety of carrier phrases. Test: Both speakers were recorded producing the test label *woogie,* in the carrier phrase *"Ooo! A woogie!"*. The child speaker was also recorded producing the correct label *book* in the carrier phrase *"Do you see the book?"*, to be used as a warm-up test trial. *Book* was chosen for this purpose because it is a highly familiar object label, and it begins with a sound that is among the very first productions children master. Audio for the warm-up and test trials were extracted from the video recordings and adjusted to be of equal perceived intensity (see Appendix Table C2 for additional information).

Training and Familiarization Trials. For the familiarization and training trials, images of the objects were digitally added to the speaker videos. A detailed description of these trials is presented in Figure 7a. Half of the training trials and half of the familiarization trials showed the labelled object in the upper left corner of the screen (before looming in the center), and half showed it in the upper right. For the two training trials, the novel object was labelled four times per trial; different carrier phrases were used for each trial. For the familiarization trials, the familiar object was labelled two times per trial; different carrier phrases were used

(a) Training and Familiarization

- speaker fades in (500 ms)
- object fades in (500 ms) in either the upper left or right corner
- video of speaker labelling object
- speaker fades out (500 ms)
- object looms to centre of screen and remains there (1,500 ms)
- object fades out (500 ms)

(b) Test



- objects appear alone for 3 seconds (baseline phase)
- speaker image fades in (100 ms) while objects remain on screen
- audio of speaker labelling object
- speaker image fades out (100 ms) while objects remain on screen
- objects appear alone to end of trial

Figure 7. Detailed description of (a) Training and Familiarization trials and (b) Test trials, with screenshots of example trials taken during object labelling.

for the two speakers.

Test Trials. For the warm-up and test trials, 2 stationary objects (both familiar for the warm-up trial; 1 trained and 1 untrained for the test trials) appeared on the left and right side of the screen and remained there throughout the trial. Each trial was 9 seconds in length. A detailed description of these trials is presented in Figure 7b.

Procedure. Toddlers were tested using the Intermodal Preferential Looking Procedure in a sound attenuated room. The child sat on the parent's lap while the parent listened to music over circumaural headphones. In front of them was a 42-inch widescreen television and two hidden speakers located at the base of the television that were connected to a computer in an adjacent room. The participants were monitored over closed-circuit video feed through a camera that was centrally located beneath the television and hidden behind a black curtain. The video feed was recorded for later off-line coding. See Figure 8 for a schematic representation of the study procedure.

Training Phase. Toddlers first saw 2 trials of the adult labeling a novel object (either blue novel or pink novel; Figure 7) with the novel label *roogie*.

Familiarization Phase. Toddlers then saw the same adult correctly labeling four familiar objects with an /ɹ/ onset (e.g., *rainbow*), followed by the child speaker labelling the same objects with an /ɹ/→[w] mispronunciation (e.g., *wainbow* for rainbow). These 2 blocks of trials (adult then child) were repeated a second time. The four familiar objects were presented in random order for each speaker.

Training Reminder Phase. Toddlers were presented with 2 more training trials of the adult labelling the novel object. Participants therefore heard the novel object labelled a total of 16 times during training.

Test Phase. The test phase began with a warm-up trial from the child speaker, in which she correctly labelled a familiar object (book) in the presence of a familiar distractor (keys).

Training (2 trials)	Adult labels training object roogie
Familiarization [a]	Adult labels familiar objects (rainbow, rooster, rattle, rabbit) Child labels familiar objects (wainbow, wooster, wattle, wabbit) *Repeat adult and child familiarization blocks*
Training Reminder	*Repeat adult training block*
Test (warm-up)	Child labels book
Test [b]	Each speaker uses label woogie (1 trial / speaker)

a. objects were presented in random order for each speaker
b. speaker order differed across participants

Figure 8. Schematic of the experimental procedure used in Experiments 4 and 5.

This was done to familiarize toddlers with the format of the test trials and to reacquaint them with the child's voice. The warm-up trial was followed by two test trials (one for each speaker). In these trials, participants saw the novel training object paired with a never-before-seen (untrained) object and heard the test label *woogie* from both speakers. For the two test trials, speaker order (child first, adult first), training object (blue novel, pink novel), and side of trained object (left, right) was counterbalanced across participants.

Following the testing session, the parent completed a questionnaire on their child's comprehension and production of the experimental items (both familiar and novel), the amount of time their child interacted with other children (siblings, daycare, and playgroups) on a weekly basis, and the number of individuals (toddlers, pre-schoolers, older children, adults) they interacted with throughout the day.

Analysis. We defined a-priori a 3-second baseline phase that occurred prior to the audio stimuli, and a 3-second naming phase. This naming phase normally would have started ~200 to 300 ms post target onset (as has been done in multiple studies of young children's word recognition to allow them time to program an eye movement). However, this was deemed inappropriate because participants took much longer to disengage from the image of the speaker following word onset. We therefore examined the warm-up trial (*book*) to determine the time point at which more than half of the participants were no longer fixating the image of the speaker. We used this time point as the onset of the naming phase for all trials. This resulted in the naming phase beginning 1100 ms after the start of the target word. This point was consistent across both experiments presented in this chapter.

Looking behavior was coded off-line using customized software at a rate of 30 frames per second (~33.33 ms/frame). For each of the two test trials, the proportion of time toddlers spent looking at the *trained* object (out of the total time spent looking at the two objects) was calculated for both baseline and naming phases. A difference score was then computed (Naming minus Baseline), which indicates how much toddlers changed their looking to the trained object from baseline to naming. A positive score indicates that they increased their looking to the trained object following naming, while a negative score indicates that they decreased their looking to that object.[6]

For a participant to be included, for each test trial the participant needed to look at each of the objects for a minimum of 6 frames (200ms)[7] during the baseline phase, as well as attend to the objects for a minimum of 1 second total during the baseline and naming phases. These criteria resulted in the exclusion of four 2-year-olds.

Data were analysed using linear mixed effects regression (Baayen et al., 2008) with R's (R Core Team, 2018) *lme4* package (Bates et al., 2015). The within-participant fixed effect of Speaker (Adult, Child) was treatment coded with the Adult Speaker as the reference condition (i.e., the model's intercept); the between-participant fixed effect of Order (Adult First, Child First) was contrast-coded. The random effects structure of all linear mixed models included only random intercepts for subject. The *lmerTest* package (Kuznetsova et al., 2017), which uses Satterthwaite's method for approximating degrees of freedom, was used to assess the significance of intercepts (Luke, 2017). The intercept of models that include the fixed effect of Speaker assess the change in looking from baseline for the Adult Speaker. To assess the change in looking for the child speaker (and the warm-up trial), separate models were used. Because these models did not include at least 2 trials per subject (the minimum

[6] The warm-up trial was analysed using the proportion of time spent looking at the *labelled* object via linear model with Age entered as a fixed factor. Overall, participants increased their looking to the book after hearing it being labelled (β_0 = 0.145, β_0 *SE* = 0.034, $t(61)$ = 4.320, p < .001, d = 0.54), with no difference across age groups (β = -0.045, β *SE* = 0.067, $t(61)$ = -0.665, p = .508, d = -0.17).

[7] Fixations last for 200 to 300 ms between saccades. The lower end of this range was used due to the object pair being the same on both test trials. The pattern of data remained the same when a more conservative criterion was used.

requirement for the inclusion of a random effect), standard linear models (that do not include random effects) were used. Random effects error variance was used as an estimate of σ^2 to calculate Cohen's *d* for model estimates (Brysbaert & Stevens, 2018; Westfall et al., 2014). Likewise, residual standard error was used as an estimate of σ in models without random effects.

Results & Discussion

We first assessed looking during the baseline phase of the two test trials against chance (0.5). To do this, baseline looking was centered around 0 by subtracting 0.5 from each value, thus making the intercept equivalent to the deviation from chance. An intercept only linear mixed effects regression model showed that 2-year-olds looked marginally more to the untrained object during baseline (β_0 = -0.041, β_0 *SE* = 0.023, *t*(31) = -1.81, *p* = .08, *d* = -0.24). This suggests a slight preference for the untrained (never-before-seen) object at test.

Our main analyses involved children's change in looking (difference scores) on the two test trials. We examined children's interpretation of the label *woogie* with a model that included Speaker (Adult, Child), Order (Adult First, Child First), and their interaction as fixed effects. Maximum likelihood comparisons revealed only a Speaker x Order interaction, $\chi^2(1)$ = 4.23, *p* = .04 (β = -0.261, β *SE* = 0.127, *d* = -0.90); there was no main effect of either Speaker or Order, $\chi^2(1) \leq 1.97$, $p \geq .160$ (Figure 9a left panel). The model intercept revealed that the 2-year-olds did not change their looking from baseline for the Adult Speaker (β_0 = 0.027, β_0 *SE* = 0.051, *t*(57) = 0.534, *p* = .595, *d* = 0.09). To assess performance from baseline for the Child Speaker, an intercept only linear model was used, revealing that the 2-year-olds did not change their looking from baseline for the Child Speaker either (β_0 = 0.068, β_0 *SE* = 0.058, *t*(31) = 1.159, *p* = .255, *d* = 0.20).

To examine the interaction, models were generated for each order with Speaker entered as a fixed effect. Model estimates reveal effects in opposing direction for each order (Adult First: β = 0.171, β *SE* = 0.107, *d* = 0.51; Child First: β = -0.090, β *SE* = 0.069, *d* = -0.38), suggesting a change across trials. A re-analysis of the data based on Trial (1, 2)[8] instead of Speaker confirms this observation, revealing that 2-year-olds did not change their looking from baseline on the first trial (β_0 = -0.018, β_0 *SE* = 0.051, $t(57) = -0.348$, $p = .729$, $d = -0.06$), and that they significantly increased their looking to the trained object from trial 1 to trial 2, $\chi^2(1) = 4.17$, $p = .041$ (β = 0.130, β *SE* = 0.063, *d* = 0.45). There was no main effect of Speaker Order or interaction with Trial, $\chi^2(1) \leq 1.97$, $p \geq .160$, indicating that regardless of Speaker, 2-year-olds increased their looking to the trained object across trials.

Overall, 2-year-olds did not treat the adult's *woogie* as a label for the untrained object, nor did they treat the child's *woogie* as a version of the training object's label. Instead they became increasingly more likely to attribute the label *woogie* to the trained object across trials, for both speakers. This suggests that children, over the course of the test, were increasingly likely to accept the /ɹ/→[w] mispronunciation for both speakers. However, given that they did not show significant mapping to either object for the adult speaker (in essence, the control speaker), it is difficult to draw any conclusions about their performance with the child speaker.

Prior work (Chapter 2; Bernier & White, 2019a) suggests that toddlers with limited experience with other children are less likely to disambiguate novel labels, that is to map novel labels to novel objects (as was expected for the adult pronunciation of *woogie*). To see whether these results were being influenced by participants with limited experience with other children, we reran the above analyses including only the 21 participants with the highest level of exposure to other children (approximately 10 participants per order). This

[8] Trial was treatment coded with Trial 1 as reference (i.e., the model intercept).

group had between 23 and 50 hours per week with other children (M = 38.2). Two-year-olds excluded from this analysis had less than 20 hours per week with other children (M = 6.7). Maximum likelihood comparisons revealed the same pattern as the larger group: there was a marginal but large Speaker x Order interaction, $\chi^2(1) = 2.99$, $p = .084$ ($\beta = -0.263$, β SE = 0.155, $d = -1.03$), and no main effect of either Speaker or Order, $\chi^2(1) \leq 0.49$, $p \geq .483$ (Figure 9a right panel). The 2-year-olds with more experience with other children also did not change their looking from baseline for either the Adult Speaker ($\beta_0 = -0.011$, β_0 SE = 0.056, $t(38) = -0.199$, $p = .843$, $d = -0.04$), or the Child Speaker ($\beta_0 = 0.041$, β_0 SE = 0.065, $t(20)$ = 0.629, $p = .537$, $d = 0.14$). Finally, re-analysis by Trial (1, 2), again shows the same pattern as the larger group: baseline looking on the first trial ($\beta_0 = -0.053$, β_0 SE = 0.056, $t(38) = -0.945$, $p = .351$, $d = -0.21$), followed by a marginal though sizeable increase in looking to the trained object from trial 1 to trial 2, $\chi^2(1) = 3.05$, $p = .081$ ($\beta = 0.131$, β SE = 0.078, $d = 0.52$).

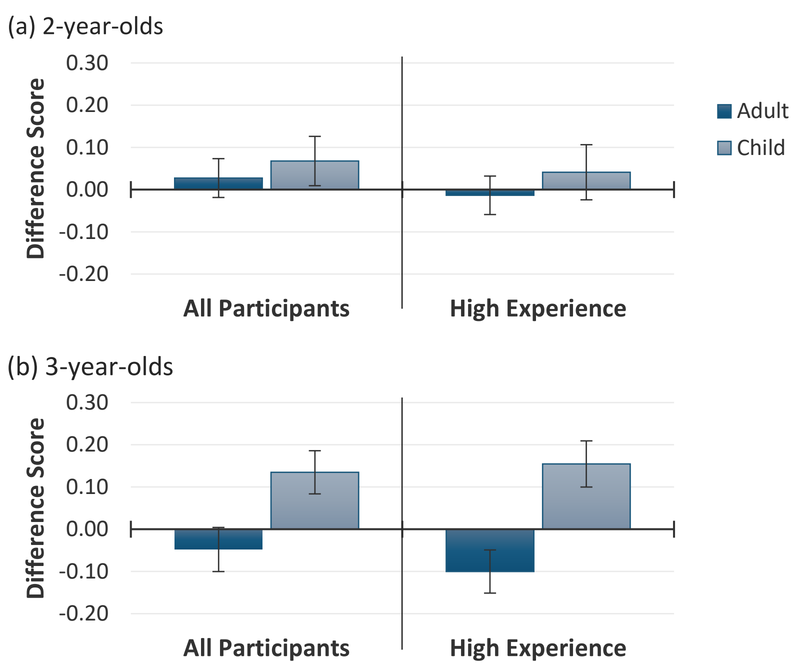


Figure 9. Estimates of children's change in looking to the trained object for (a) 2-year-olds (Experiment 4) and (b) 3-year-olds (Experiment 5). The left side shows the results for all participants, the right side the results for the participants with the most exposure to other children (> 20 hours/week for 2-year-olds, > 10 hours/week for 3-year-olds excluding one outlier). Error bars represent standard error.

Overall, neither toddlers as a group, nor those with more experience with other children, appear to undo the child's /ɹ/→[w] mispronunciation with the novel label. Instead, 2-year-olds appear to look more to the trained object across trials (regardless of which speaker they heard first). We will come back to this in the General Discussion. For now, we turn to slightly older children, who are more likely to demonstrate a strong disambiguation response for the adult's novel label *woogie*, thereby allowing us to determine whether their response to the child's label demonstrates that they have undone the child's mispronunciation.

Experiment 5: Expectations in 3-year-olds

Experiment 5 tested 3-year-olds for two reasons. First, 3-year-olds are more likely to demonstrate a disambiguation response for the adult's test label, meaning that their response to the child's test label is more likely to be interpretable. Second, 3-year-olds have more experience with other children, and with these sounds in particular, than 2-year-olds, meaning that they are more likely to have expectations about the way children speak. As predicted for the 2-year-olds, if 3-year-olds can undo the child's /ɹ/→[w] mispronunciation, they should map the child's *woogie* to the trained object, but map the adult's *woogie* to the untrained (never-before-seen) object. As with Experiment 4, we also considered participants' experience with other child speakers due to its potential role in the development of these kinds of expectations.

Method

Participants. A total of 32 monolingual English-learning 3-year-olds (mean age 38.1 months; range 36.3-40.0) were recruited from the Waterloo Region of Ontario, Canada. Five additional 3-year-olds were tested but not included due to not meeting trial inclusion criteria (see Experiment 4 Methods) on one of the test trials.

Stimuli, Procedure, and Analysis. Same as Experiment 4.

Results & Discussion

As with Experiment 4, we first assessed looking during the baseline phase against chance (0.5), by centering baseline looking around 0. An intercept only linear mixed effects regression model showed that 3-year-olds looked equally at the trained and untrained object during baseline (β_0 = -0.028, β_0 *SE* = 0.029, *t*(30) = -0.75 *p* = .337, *d* = -0.14).

Our main analyses involved children's change in looking (difference scores) on the two test trials. We examined children's interpretation of the label *woogie* with a model that included Speaker (Adult, Child), Order (Adult First, Child First), and their interaction as fixed effects. Maximum likelihood comparisons revealed a main effect of Speaker, $\chi^2(1) = 6.14$, *p* = .013 (β = 0.183, β *SE* = 0.072, *d* = 0.65), and no main effect of Order or interaction, $\chi^2(1) \leq 0.85$, $p \geq .356$ (Figure 9b left panel). Therefore, 3-year-olds, unlike 2-year-olds, showed a different pattern of performance for the two speakers. The model intercept revealed that 3-year-olds, like 2-year-olds, did not change their looking from baseline for the Adult Speaker (β_0 = -0.048, β_0 *SE* = 0.052, *t*(58) = -0.948, *p* = .347, *d* = -0.17). However, an intercept only linear model revealed that they significantly increased their looking to the trained object from baseline for the Child Speaker (β_0 = 0.135, β_0 *SE* = 0.049, *t*(30) = 2.718, *p* = .011, *d* = 0.49).

Overall, 3-year-olds appear to treat the child's label *woogie* as a version of the training object label. However, just like the 2-year-olds, they did not show a strong disambiguation response for the adult's label. As with the 2-year-olds in Experiment 4, we re-ran the above analyses including only the 20 participants with the highest level of exposure to other children (10 participants per order). This group had between 12 and 45 hours per week with other children (*M* = 30.3). Those excluded from this analysis had 10 hours or less per week with other children (*M* = 3.7). Maximum likelihood comparisons revealed the same pattern as the larger group: there was a significant effect of Speaker, $\chi^2(1) = 6.78$, $p = .009$ ($\beta = 0.207$, β $SE = 0.080$, $d = 0.82$) with no main effect of Order or interaction, $\chi^2(1) \leq 0.45$, $p \geq .503$. Although the model showed that they trended towards looking to the untrained object for the Adult Speaker, this difference was not significant ($\beta_0 = -0.072$, β_0 $SE = 0.057$, $t(36) = -1.270$, $p = .212$, $d = -0.28$). An intercept only linear model showed that they significantly increased their looking to the trained object from baseline for the Child Speaker ($\beta_0 = 0.136$, β_0 $SE = 0.056$, $t(19) = 2.455$, $p = .024$, $d = 0.55$). However, one 3-year-old with High Experience had a score for the adult speaker that was more than 2 *SD* above the mean. Excluding this one participant resulted in a larger main effect of Speaker, $\chi^2(1) = 11.18$, $p < .001$ ($\beta = 0.255$, β $SE = 0.071$, $d = 1.09$), accompanied by a marginal decrease in looking for the Adult Speaker ($\beta_0 = -0.102$, β_0 $SE = 0.054$ $t(33.3) = -1.889$, $p = .068$, $d = -0.43$), and a significant increase in looking for the Child Speaker ($\beta_0 = 0.155$, β_0 $SE = 0.055$, $t(18) = 2.828$, $p = .011$, $d = 0.65$; Figure 9b right panel).

Overall, the results for the 3-year-olds, who treated the child and adult pronunciations of *woogie* differently, are in contrast to those of the 2-year-olds, who did not, suggesting a developmental shift across these two age groups. This difference was explored with a model that included Age (2yr, 3yr) as well as Speaker and Order as fixed factors. Maximum likelihood comparisons revealed a significant 3-way interaction, $\chi^2(1) = 4.21$, $p = .040$ ($\beta = 0.392$, β $SE = 0.194$, $d = 1.38$), confirming the observed difference.[9] There were no other effects of Age, $\chi^2(1) \leq 2.04$, $p \geq .154$. Thus, the two age groups performed differently in this task.

[9] An analysis of the 3-year-olds by Trial showed no main effect of Trial, $\chi^2(1) = 0.67$, $p = .414$ ($\beta = -0.066$, β $SE = 0.071$, $d = -0.23$), but did reveal a Trial x Speaker Order interaction, $\chi^2(1) = 6.67$, $p = .010$ ($\beta = -0.367$, β $SE = 0.143$, $d = -1.30$) which indicates differential treatment by Speaker.

Chapter Discussion: Experiments 4 and 5

Across two experiments, we tested young children's inferences about the intended meaning of a novel label from an adult and from a child who was shown to mispronounce familiar words. For the adult speaker, we expected that the novel label would be interpreted as a label for a new, never-before-seen, object, because of the commonly observed disambiguation response in this age group (e.g., Bion et al., 2013). However, for the child speaker, if children either had pre-existing expectations about common mispronunciations or learned the child's pronunciation patterns during the experiment, the novel label should have been interpreted as a mispronunciation of the newly learned word. We found that 2-year-olds did not interpret the two labels in this way, but that at least some 3-year-olds did. Specifically, the 3-year-olds showed a different pattern of performance for the two speakers, with higher looking to the never-before-seen object for the adult speaker and significantly higher looking to the trained object for the child speaker. Only 3-year-olds' performance for the adult speaker (but not the child speaker) was influenced by experience with other children.

This suggests that by 3 years, at least some children may have expectations about the way children speak. However, this conclusion comes with an important caveat. Children in this study were exposed to the child speaker mispronouncing familiar labels in the same way as the test label. It is therefore possible that they generalized from the familiar words to the novel word without the need for a separate expectation for child speakers that contains the /ɹ/→[w] transformation. One way to determine whether 3-year-olds were using expectations (as opposed to generalizing from familiarization), would be to test them on the mispronounced label without providing them with examples of how the child mispronounces the familiar words. If they have not previously heard the child pronounce labels such as *wainbow*, then interpreting the child's label *woogie* as a (mispronounced) label for the trained object would mean that they have relied on prior experience with child pronunciations to infer this mapping. This study is currently in progress. Another way to determine whether 3-year-olds have these kinds of expectations would be to use a test mispronunciation that is unlikely to occur in children's speech (e.g., a /w/→[ɹ]). If they have expectations about children's speech, then undoing a mispronunciation that conflicts with their expectations should be more difficult.

Turning to the 2-year-olds, their results suggest that they did not know how to interpret either the child's or adult's label. One possible reason is that this contrast was too subtle for them to perceive at this stage of phonological development. Even 4-year-olds' perception of /ɹ/ has been found to be less accurate than their perception of more robust and earlier acquired sounds, like /k/ (Hearnshaw, Baker, & Munro, 2018). However, given that previous work in this dissertation (Chapter 3; Bernier & White, 2019b) has shown that the /ɹ/→[w] mispronunciation is detected in familiar words by 21- to 25-month-olds, it is unlikely that the 2-year-olds had difficulty detecting the mispronunciation during familiarization. Nevertheless, it remains possible that 2-year-olds have difficulty detecting the /ɹ/→[w] mismatch when it occurs in unfamiliar words, as was the case in the present test phase.

Consideration of children's early word knowledge supports the contention that /ɹ/ is a weakly represented sound. A summary of WordBank's American English-learning children shows that 25- to 30-month-olds (n = 1,999) do produce words that begin with /ɹ/ as much as they do words with other onsets (Frank et al., 2017), indicating that they are not unfamiliar with the specific sound used in this study. However, children's ability to detect changes to new words relies on the robustness of their representations of the individual sounds. The more robustly a sound is represented, the more stable and better recognized it will be in new contexts. It has been argued that sounds that occur in more words are likely to be more robustly represented than sounds heard in only a few words (Edwards et al., 2015). And in fact, the /ɹ/ sound used in this study occurs at the beginning (the most salient position) of only 14 words of the MacArthur Bates Communicative Development Inventory: Words and Sentences (Fenson et al., 1994). This is in contrast to 58 words (excluding onset clusters) that begin with /b/ and 44 that begin with /p/ – the two most common onsets. This supports the idea that young children are likely have less robust representations of /ɹ/ than more common onsets, which may be contributing to 2-year-olds' difficulty in the present study. It is therefore possible that 2-year-olds may show better performance for the adult speaker (and perhaps even demonstrate expectations about child mispronunciations) if they are presented with a more well-known onset or a type of error that is more easily detected (e.g., stopping – the most easily detected mispronunciation in chapter 3).

In summary, 3-year-olds showed differential responding for child and adult speakers, consistent with the possibility that they have different expectations for these two types of speakers. However, this may have been due to them generalizing from familiarization to test. And although 2-year-olds did not perform as expected for the adult speaker, they may have had difficulty processing the specific contrast used in this study. It therefore remains possible that 3-year-olds (and maybe 2-year-olds), have expectations about the kinds of speech errors children typically make. If so, this would be in line with the results of Krueger et al. (2018) who found that 5-year-olds were more tolerant of common mispronunciations than rare ones from adult speakers. However, their selection task data conflict with other looking-time data (Chapter 3; Bernier & White, 2019b; Krueger et al., 2018) demonstrating that children are more tolerant of rare or infrequent mispronunciations from adult speakers, and that they are equally tolerant of common and rare mispronunciations from a child speaker.

In conclusion, having expectations about groups of individuals helps us efficiently process the speech of unfamiliar individuals, including children. By 3 years of age (and possibly earlier), children may be able to make inferences about the way a child is likely to produce a new word based on limited information. Future work is needed to further refine our understanding of the extent and development of these expectations.

Chapter Five: General Discussion

This dissertation examined toddlers' perception of other children's speech with two primary objectives: (1) to determine how well toddlers process the speech of a slightly older child, and (2) to determine whether toddlers have expectations about what other children sound like – either generally as speakers who mispronounce words, or more specifically as speakers who make certain kinds of errors. A secondary objective of this dissertation was to examine a potential role of experience in toddlers' processing of child speech. Below I will discuss the results in light of each of these objectives in turn.

Toddlers' Processing of Older Children's Speech

The work in Chapters 2 and 3 demonstrates that 20- to 24-month-olds are able to process the speech of a first-grade child as well as the speech of an adult. Not only were toddlers in Chapter 2 able to recognize familiar words produced by a 7-year-old speaker as accurately as those from an adult, they were just as sensitive to phonetic changes in the child's speech as they have been in previous work with adult speech (e.g., White & Morgan, 2008). This sensitivity was also observed in Chapter 3 when their processing of a 6-year-old's speech was compared with a different adult speaker. In that study, toddlers were, overall, equally sensitive to mispronunciations in child and adult speech. That these two series of studies used different child and adult speakers suggests that toddlers' accuracy for child speech is not specific to any one speaker, but is something more broadly true about their processing of speech from typically developing school-aged children.

These results show that, despite the variability present in the speech of young school-aged children (e.g., Koenig, Lucero, & Perlman, 2008; Lee, Potamianos, & Narayanan, 1999; Nissen & Fox, 2005), toddlers easily understand the speech of children as young as 6 years old (in contrast to the speech of much younger children; Cooper, Fecher, & Johnson, 2018; Dodd, 1975). However, it remains possible that under more challenging listening conditions, such as with background noise, toddlers may find the speech of other children more difficult to process. This kind of decrement has been shown to occur for accented speech in both adults and children, such that there is a bigger disadvantage in the processing of accented speech (vs. familiar speech) in noisy relative to quiet conditions (Bent & Holt, 2018). Toddlers have also been shown to perform poorly in word recognition tasks under noisy conditions with native speakers (Newman, 2011). Together, this suggests that background noise could reveal processing differences between child and adult speech. Even so, this suggests that the speech of early school-age children, who are already skillful (though immature) language users (Tolchinsky, 2004), may constitute input that toddlers can use for their own language learning (more discussion on this below).

Toddlers' Expectations About the Way Children Speak

The work in Chapter 2 examined indirectly whether 20- to 24-month-olds have expectations about the way other children speak, by asking about their tolerance of mispronunciations in the speech of a 7-year-old. This study found that toddlers were not tolerant of the child speaker's mispronunciations (i.e., they showed graded sensitivity rather than a general acceptance of the mispronunciations). This result suggests that toddlers do not expect children to produce speech errors (or if they do, they still do not tolerate them). However, the mispronunciations used in this study were not all equally likely to be produced by toddlers; some are common in children's speech, others rarely or never observed. Chapters 3 and 4 more directly assessed toddlers' expectations about child speech by examining contrasts that are (or are not) typically mispronounced by young children.

In Chapter 3, 20- to 24-month-olds appeared to be more tolerant of common childhood mispronunciations (i.e., showed a smaller mispronunciation penalty) when they were produced by a child compared to an adult, as one would predict if toddlers expect children to make errors and are thus more tolerant of their mispronunciations. However, this difference was not significant, so must be interpreted with caution. Moreover, contrary to a pattern of greater tolerance for child speech, when toddlers were presented with infrequent mispronunciations, they showed a negligible mispronunciation penalty for the adult speaker, but a moderately sized penalty for the child speaker. Because it is highly unlikely that toddlers detect only mispronunciations produced by a child (but not an adult), or expect that only adults would mispronounce an object label (but not a child), this result must be driven by something other than pure perceptual sensitivity or expectations.

But do toddlers expect children to make specific errors? In Chapter 3, 20- to 24-month-olds were found to treat the mispronunciations of a 6-year-old with the same level of sensitivity regardless of whether the mispronunciations were common in children's speech (e.g., /ɹ/→[w]) or rare (e.g., /w/→[ɹ]). In Chapter 4, a developmental difference was found in 2- and 3-year-olds' ability to process a 5-year-old's common /ɹ/→[w] mispronunciation of a newly learned word. Although 2-year-olds had difficulty with this contrast when it was presented in a new word, at least some 3-year-olds inferred that the child was referring to the trained object with the mispronounced label. However, the nature of the familiarization makes it unclear whether the 3-year-olds were indeed relying on expectations or whether they were generalizing from their recent exposure to the child mispronouncing familiar words in this same way.

In total, these results suggest that children under 3 years do not expect children to produce specific types of errors or even to produce speech errors more generally. One possible reason for this is that toddlers simply do not have expectations about the way other children speak. This would be consistent with the findings of Yurovsky et al. (2017). In that study, although 4-5-year-olds developed and used speaker-specific expectations for an adult speaker, 3-year-olds did not. However, an inability to build expectations about groups of speakers at 3 years of age would be inconsistent with Weatherhead and White (2018), who found that 16-month-olds expected speakers who looked familiar (i.e., had a similarly fair complexion) to produce words with a familiar pronunciation, but had no such expectation for speakers who looked different (i.e., had a darker complexion). There are at least three possible reasons that toddlers did not demonstrate any evidence for expectations here. First, age may not be a category that toddlers initially link to language variation. In other words, they may expect children to speak like other members of their language community. Second, it is possible that toddlers do not expect older children to produce errors any more than they do adult speakers, but that they do have expectations for younger children whose speech contains more overt errors than that of older children. And, third, it is possible that the sentence frames leading up to the target words in these studies provided toddlers with cues as to whether the speaker was likely to mispronounce the target word. School-age children do not produce many overt errors; therefore, the accuracy of the speech produced prior to the target word may have led toddlers to expect adult-like productions. In contrast, they may have different expectations about speakers who do make grammatical or phonetic errors. This is something that pre-school children are sensitive to, as they appear to expect accented speakers, but not native speakers, to make grammatical errors (Rett & White, 2019). In summary, it remains unclear from these results whether young children have expectations but do not use them for 5- to 7-year-old speakers, or whether they have not yet developed any expectations at all.

The Role of Experience

A secondary objective of this dissertation was to assess the potential role of experience in toddlers' processing of child speech. For this dissertation, experience was determined by the amount of time toddlers spent with other children in group activities outside the home (daycare, informal playgroups, organized community activities). Although this measure was used for each experiment, the cutoffs used to determine High vs Low experience differed somewhat for each experiment (Table 10). A median split was used for Experiments 1 and 2. However, in order to get a cleaner distinction between High and Low experience in Experiment 3, only the participants with the highest and lowest amounts of experience were assessed (the middle 20 were excluded from analysis). Finally, in Experiments 4 and 5, participants with the least amount of experience were excluded to determine whether difficulty with disambiguation may have been influencing the results, particularly for the adult speaker.

In Chapter 2, a subtle difference was found in toddlers' sensitivity to phonetic changes as a function of experience. Toddlers with more experience with other children showed a larger mispronunciation penalty for both one- and three-feature changes compared to toddlers with limited experience. This result suggests that hearing a variety of other children's immature speech has an impact on how toddlers process a child's speech (in particular, by making them

Table 10. Hours per week participants in the High and Low experience groups from each experiment spent with children other than siblings

	High Experience		Low Experience	
	N	Total hours [*M* (range)]	N	Total hours [*M* (range)]
Experiment 1	24	29.9 (10-48)	24	2.1 (0-8)
Experiment 2	20	30.6 (10-50)	20	3.1 (0-7)
Experiment 3	50	41.6 (35-56)	50	3.3 (0-10)
Experiment 4	21	38.2 (23-50)	11	6.7 (0-19)
Experiment 5	20	20.3 (12-45)	10	3.7 (0-10)

more sensitive). However, the difference between experience groups was not significant.

In Chapter 3, where child and adult speech were directly compared, processing differences were indeed found across speakers. Although toddlers' processing of adult speech did not differ as a function of experience, toddlers' processing of child speech differed subtly. Toddlers with more experience with other children showed a smaller penalty for common mispronunciations than toddlers with less experience. No such difference was found for infrequent mispronunciations. However, upon closer inspection, the primary difference for the common mispronunciations appears to be in toddlers' treatment of correctly pronounced labels rather than mispronounced labels. In other words, the smaller mispronunciation effect was due to lower looking for correct labels, rather than higher looking for mispronounced labels. In fact, the High experience group showed the same change from baseline for commonly and infrequently mispronounced labels. Overall, these null effects suggest that toddlers process child and adult speech similarly regardless of experience.

In Chapter 4, toddlers' processing of child and adult speech was again directly compared. However, this time it was examined in the context of a word learning task with a /ɹ/→[w] error. Here, experience with other children did not influence 2-year-olds' processing, but it did somewhat influence 3-year-olds. However, rather than influencing 3-year-olds' processing of the child's mispronounced label, experience with other children instead subtly influenced the processing of the adult's label. In other words, 3-year-olds with more experience (compared to the group as a whole) were more likely to map the adult's mispronounced label to the novel object. In total, effects of experience with other children (when they are found) are small and thus do not appear to be a significant influence on toddlers' processing of children's speech.

However, an influence of experience on toddlers' *novel* label processing was observed to some degree across all of the experiments in this dissertation. In Chapter 2, where this difference was first observed, only toddlers with more experience showed the expected response of mapping novel labels to novel objects. This effect of experience held across both child and adult speakers, suggesting a general difference in the processing of novel labels. This effect was again found in Chapter 3, only this time it was restricted to the child speaker. And finally, an experience effect was observed in Chapter 4 as a subtle difference in 3-year-olds' willingness to map an adult's mispronounced label to the novel object. However, it is unclear what is driving this difference in novel label mapping. An additional survey (Chapter 2) of the same population did not show any specific difference across the High and Low Experience groups as we defined them that might account for this difference (e.g., vocabulary size, maternal education). Together, these findings suggest that there is some aspect of toddlers' experience with groups of children that is influencing their interpretation of novel labels.

One possibility is that the experience effect is due to an increased amount of input from different adults rather than children. In daycares and other group settings, the number of adults increases as well as the number of children. Thus it is likely that toddlers who engage in more group activities (whether in daycare or elsewhere) are also engaging with more adults. It is therefore possible that more interactions with a greater variety of adults, which influences toddlers' processing abilities more broadly (e.g., Rost & McMurray, 2009), is the source of this effect. Given other demonstrations that language experience affects children's interpretation of novel labels (i.e., bilinguals who have more words in both languages for the same concepts are less likely to exhibit a disambiguation response; Byers-Heinlein & Werker, 2013), this difference is an intriguing one to follow up on in future research.

Re-Defining Experience

Across the studies of this dissertation, toddlers showed minimal (if any) effects of experience or expectations on processing the properties of child speech. One possibility for this is that experience with other children does not have much effect over and beyond the effect of hearing their own speech. However, another possibility is that the measure of experience used here is imprecise and does not clearly capture the type of experience that matters. A broad measure such as the one used in these studies by necessity encompasses a wide variety of experiences, with some of these experiences possibly being more meaningful for language development than others.

One such factor that might influence language development is the type of play a toddler typically engages in when they are with other children. Solitary, onlooker, and parallel play occur in toddlers to varying degrees, as does the possibility of associative play when toddlers are included in older children's play. Each of these types of play requires a different level of interaction between the children involved. Although toddlers engaged in parallel and onlooker play will be attending to those around them, toddlers engaged in solitary play will not. Studies of parental input show that parent-child interactions (conversational turns) are important for drawing the child's attention to language, thus supporting language learning (e.g., Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015; Romeo et al., 2018). Therefore, as with parent-child interactions, situations where one child is attending to the speech of another child will likely provide the child with more optimal language learning opportunities. It is therefore possible that the frequency with which toddlers engage in solitary versus other forms of play may be an important mediator of whether their experience with other children influences their language development. For example, a toddler engaged in parallel or associative play with preschoolers is likely to gain more experience with child speech (due to the desire to understand the other children) than the toddler engaged in solitary play next to them (who is only passively hearing the children speak).

Along the same lines, the age of the other children may be another important factor (e.g., other toddlers, vs preschoolers or older children). Children do not typically engage in associative play (which involves social interaction among children) until after 3 years. So, although toddlers are unlikely to spontaneously engage in this type of play, toddlers who are interacting with older children may find themselves included in associative play. In addition to this, toddlers' difficulty processing the speech of other toddlers (Cooper et al., 2018; Dodd, 1975) suggests that toddlers' speech will be less important as a source of experience than the speech of older children. Overall, toddlers who are frequently included in pre-schoolers' associative play will likely gain more experience with child speech, compared to toddlers who typically interact only with other toddlers (who are less likely to direct speech to them and whose speech is more challenging to process).

In a recent study, Havron et al. (2019) did not find an effect of the size of age gap between siblings on younger children's language outcomes. However, the target children in this study ranged from 2- to 6-years, therefore it remains possible that the degree of age gap between interacting children may be stronger for younger children's language development. For example, a 1 versus 3 year age gap between a 16-month-old learner and older children will mean a larger difference in the quality of the input from the older children compared to this same age gap for a 5-year-old learner.

In addition to age, the gender of the other children may also be important. The above-mentioned study (Havron et al., 2019) found that children with older sisters had better language skills than children with older brothers. The authors suggest that one possible reason for this result is that girls are more likely to engage younger siblings in more complex verbal communication than boys. Therefore, toddlers who are frequently included in the play of older girls may gain more experience with child speech (due to the increased likelihood of verbal interaction), than toddlers who are frequently included in the play of older boys.

Assessing Younger Listeners and Speakers

Although this dissertation found that children at least 20 months old were able to process the speech of 5- to 7-year-old children as well as speech of adults, it remains possible that younger children may have difficulty with this same type of child speech. Work with accented speech has shown that children younger than those assessed in the present studies have difficulty accommodating accented speech, even after exposure (e.g., Schmale, Cristia, & Seidl, 2012; van Heugten & Johnson, 2014; White & Aslin, 2011). It is therefore possible that infants (e.g., 16 months) would have difficulty accommodating the variation in the speech of school-aged children. If so, stronger effects of experience might be observed in this age group. Further investigation is needed to determine whether this is the case.

Additionally, the speech assessed in this dissertation was that of 5- to 7-year-olds. Effects of experience may be more obvious with younger children's speech. Although toddlers have in general been shown to have difficulty with the speech of other toddlers (Cooper et al., 2018; Dodd, 1975), individual differences in experience have not been explored. That said, at this age there are often multiple deviations per word (e.g., *beya* [bejʌ] for umbrella), which may make the mapping too difficult, regardless of how much experience a toddler has listening to child speech. In contrast, 3- to 4-year-olds have fewer deviations per word than younger children (e.g., they are more likely to say something like *umbella* [ʌmbelʌ] for umbrella), but are still considerably less accurate than 5- to 7-year-olds. Therefore, stronger effects of experience might be observed when toddlers listen to the speech of 3- to 4-year-olds.

Concluding Comments

This dissertation is the first demonstration that speech from young school-age-children is highly understandable to toddlers. Despite the variation present in the speech of early school-aged children, toddlers were able to accommodate this variation. This suggests that, although typically overlooked, child speech may be a source of input for young language learners. Indeed, the Havron et al. (2019) results discussed above are consistent with this possibility.

However, in one study that directly examined the role of input from adults and children in a small Mayan Yucatec community, Shneidman and Goldin-Meadow (2012) found that toddlers' vocabulary was predicted by the speech directed at them from adults, not the speech from the children they spent most of their day with. However, it would be premature to conclude that child input is not important on the basis of a single study. Moreover, word learning may not be the best place to look for effects of child speech. Children commonly mislabel objects (e.g., calling a cat a dog), and even when they do label an object appropriately, they often mispronounce it (e.g., pronouncing cat at *tat*). Given that learning word meanings relies on an assumption that the speaker is a competent member of your language community (Clark, 2014), it is possible that toddlers do not consider other children to be good sources for learning word-object mappings (Jaswal & Neely, 2006). However, language development involves more than word learning. It also includes learning about the sound system (e.g., speech categories, phonotactic patterns), building more robust word representations and becoming efficient at word recognition, and learning about grammatical patterns and other aspects of language use. Therefore, even if child input proves to be less important for word learning, child speech could still be an important contributor to other aspects of language development.

Children do eventually learn new words from their peers. So it is not really a question of *if* children learn words from other children, but *when* they begin to do so and *from whom* they acquire them (i.e., the ages of the children involved). Although the results of this dissertation suggest that toddlers could use the speech of young children who are 3 to 5 years older as source of language input, many questions still remain. In short, more work is still needed to elucidate the role of child speech in young children's language development.

References

Altvater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua*, *120*, 1898–1909. https://doi.org/10.1016/j.lingua.2010.02.010

Altvater-Mackensen, N., van der Feest, S. V. H., & Fikkert, P. (2014). Asymmetries in early word recognition: The case of stops and fricatives. *Language Learning and Development*, *10*, 149–178. https://doi.org/10.1080/15475441.2013.808954

Amir, O., Engel, M., Shabtai, E., & Amir, N. (2012). Identification of children's gender and age by listeners. *Journal of Voice*, *26*, 313–321. https://doi.org/10.1016/j.jvoice.2011.06.001

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A Statistical basis for speech sound discrimination. *Language and Speech*, *46*, 155–182. https://doi.org/doi.org/10.1177/00238309030460020601

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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

Bent, T., & Holt, R. F. (2018). Shhh… I need quiet! Children's understanding of American, British, and Japanese-accented English speakers. *Language and Speech*, *61*, 657–673. https://doi.org/10.1177/0023830918754598

Bernard, A. (2018). *Insights on Canadian society: Results from the 2016 Census: Work activity of families with children in Canada.* Retrieved from https://www150.statcan.gc.ca/n1/pub/75-006-x/2018001/article/54969-eng.htm

Bernier, D. E., & Soderstrom, M. (2016). *[LENA recording study contrasting different child care environments]*. Unpublished raw data.

REFERENCES

Bernier, D. E., & White, K. S. (2019a). Toddlers' sensitivity to phonetic detail in child speech. *Journal of Experimental Child Psychology*, *185*, 128–147. https://doi.org/10.1016/j.jecp.2019.04.021

Bernier, D. E., & White, K. S. (2019b). Toddlers process common and infrequent childhood mispronunciations differently for child and adult speakers. *Journal of Speech, Language, and Hearing Research*, *62*, 4137-4149. https://doi.org/10.1044/2019_JSLHR-H-18-0465

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, *20*(5), 539–542. https://doi.org/https://doi.org/10.1111/j.1467-9280.2009.02327.x

Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53. https://doi.org/10.1016/j.cognition.2012.08.008

Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer*. Retrieved from http://www.praat.org/

Booth, C. L., Clarke-Stewart, K. A., Vandell, D. L., McCartney, K., & Owen, M. T. (2002). Child-care usage and mother-infant "quality time." *Journal of Marriage and Family*, *64*, 16–26. https://doi.org/10.1111/j.1741-3737.2002.00016.x

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Burchinal, M. R., Roberts, J. E., Nabors, L. A., & Bryant, D. M. (1996). Quality of center child care and infant cognitive and language development. *Child Development*, *67*, 606–620. https://doi.org/10.2307/1131835

REFERENCES

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, *12*(5), 815–823. https://doi.org/10.1111/j.1467-7687.2009.00902.x

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*, 407–416. https://doi.org/10.1016/j.cognition.2013.05.010

Cerrato, L., Falcone, M., & Paoloni, A. (2000). Subjective age estimation of telephonic voices. *Speech Communication*, *31*, 107–112. https://doi.org/1016/S0167-6393(99)00071-0

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children.pdf. *Journal of Child Language*, *17*, 205–215. https://doi.org/10.1017/S0305000900013180

Clark, E. V. (2014). Two pragmatic principles in language use and acquisition. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 105–120). https://doi.org/10.1075/tilar.10.07cla

Cooper, A., Fecher, N., & Johnson, E. K. (2018). Toddlers' comprehension of adult and child talkers: Adult targets versus vocal tract similarity. *Cognition*, *173*, 16–20. https://doi.org/10.1016/j.cognition.2017.12.013

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior & Development*, *34*, 590–601. https://doi.org/10.1016/j.infbeh.2011.06.005

Dodd, B. (1975). Children's understanding of their own phonological forms. *Quarterly Journal of Experimental Psychology*, *27*, 165–172. https://doi.org/10.1080/14640747508400477

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics & Phonetics*, *17*, 617–643. https://doi.org/10.1080/0269920031000111348

Edwards, J., Beckman, M. E., & Munson, B. (2015). Frequency effects in phonological acquisition. *Journal of Child Language*, *42*, 306–311. https://doi.org/10.1017/S0305000914000634

Falenchuk, O., Perlman, M., McMullen, E., Fletcher, B., & Shah, P. S. (2017). Education of staff in preschool aged classrooms in child care centers and child outcomes: A meta-analysis and systematic review. *PloS One*, *12*, 1–33. https://doi.org/10.1371/journal.pone.0183673

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*, 1–185. Retrieved from http://www.jstor.org/stable/1166093

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, *21*, 95–116.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*, 228–231. https://doi.org/10.1111/1467-9280.00044

Flipsen, P. Jr. (2006). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics*, *20*, 303–312. https://doi.org/10.1080/02699200400024863

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*, 677–694. https://doi.org/10.1017/S0305000916000209

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*, 8051–8056. https://doi.org/10.1073/pnas.1216438110

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, *28*, 703–712. https://doi.org/10.1177/0956797617690277

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*, 339–344. https://doi.org/10.1177/0963721415595345

Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, *6*, 1200. https://doi.org/10.3389/fpsyg.2015.01200

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*, B23–B34. https://doi.org/10.1016/S0

Hanulíková, A., van Alphen, P. M., Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*, 878–887. https://doi.org/10.1162/jocn_a_00103

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brooks Publishing.

Havron, N., Ramus, F., Heude, B., Forhan, A., Cristia, A., Peyre, H., & the EDEN Mother-Child Cohort Study Group. (2019). The effect of older siblings on language development as a function of age difference and sex. *Psychological Science*, *30*, 1333–1343. https://doi.org/10.1177/0956797619861436

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*, 865–892. https://doi.org/doi: 10.1515/ling.2010.027

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*, 351–379. https://doi.org/10.1515/TLR.2006.014

Hayden, R. E. (1950). The relative frequency of phonemes in General-American English. *Word*, *6*, 217–223. https://doi.org/10.1080/00437956.1950.11659381

Hearnshaw, S., Baker, E., & Munro, N. (2018). The speech perception skills of children with and without speech sound disorder. *Journal of Communication Disorders*, *71*, 61–71. https://doi.org/10.1016/j.jcomdis.2017.12.004

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*, 1661–1086. https://doi.org/10.1121/1.1337959

Hodson, B. W., & Paden, E. P. (1981). Phonological processes which characterize unintelligible and intelligible speech in early childhood. *Journal of Speech and Hearing Disorders*, *46*, 369. https://doi.org/10.1044/jshd.4604.369

Hoff, E. (2006). Language Experience and Language Milestones During Early Childhood. In D. McCartney, Kathleen Phillips (Ed.), *Blackwell Handbook of Early Childhood Development*. https://doi.org/10.1002/9780470757703.ch12

Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language*, *33*, 227–249. https://doi.org/10.1017/S0305000906007896

Ingram, D. (2008). Cross-Linguistic Phonological Acquisition. In M. J. Ball, M. R. Perkins, N. Müller, & S. Howard (Eds.), *The Handbook of Clinical Linguistics*. https://doi.org/10.1002/9781444301007.ch38

Jaswal, V. K., & Neely, L. A. (2006). Adults don' t always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*(9), 757–758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, *37*, 295–311. https://doi.org/10.1006/jmla.1997.2522

Klein, H. B., McAllister Byun, T., Davidson, L., Grigos, M. I., Byun, T. M., Davidson, L., … Grigos, M. I. (2013). A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech Language Pathology*, *22*, 540–533. https://doi.org/10.1044/1058-0360(2013/12-0137)

Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *The Journal of the Acoustical Society of America*, *124*, 3158–3170. https://doi.org/10.1121/1.2981639

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, *121*, 459–465. https://doi.org/10.1016/j.cognition.2011.08.015

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*, 332–338. https://doi.org/10.1111/j.1467-9280.2008.02090.x

Krueger, B. I., Storkel, H. L., & Minai, U. (2018). The influence of misarticulations on children's word identification and processing. *Journal of Speech Language and Hearing Research*, *61*, 820–836. https://doi.org/10.1044/2017_JSLHR-S-16-0379

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

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*, 44–59. https://doi.org/10.1016/j.wocn.2010.01.002

Laing, C., & Bergelson, E. (2019). Mothers' work status and 17-month-olds' productive vocabulary. *Infancy*, *24*, 101–109. https://doi.org/10.1111/infa.12265

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(97), 1455–1511. https://doi.org/10.1121/1.412064

Lev-Ari, S. (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology*, *5*, 1546. https://doi.org/10.3389/fpsyg.2014.01546

Lin, S., Inkelas, S., McDonnaughey, L., & Dohn, M. (2016). /l/ in clusters: An articulatory-acoustic study of children's productions. *Talk Presented at The 15th Conference on Laboratory Phonology*. Ithaca, NY.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Maas, E., & Mailend, M. L. (2017). Fricative contrast and coarticulation in children with and without speech sound disorders. *American Journal of Speech-Language Pathology*, *26*, 649–663. https://doi.org/10.1044/2017_AJSLP-16-0110

Magloughlin, L. (2016). Accounting for variability in North American English /ɹ/: Evidence from children's articulation. *Journal of Phonetics*, *54*, 51–67. https://doi.org/10.1016/j.wocn.2015.07.007

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, *10*, 179–204. https://doi.org/10.1080/15475441.2013.829740

Mani, N., & Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language*, *38*, 606–627. https://doi.org/10.1017/S0305000910000243

McCartney, K. (1984). Effect of quality of day care environment on children's language development. *Developmental Psychology*, *20*, 244–260. https://doi.org/10.1037/0012-1649.20.2.244

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, *131*, 509–517. https://doi.org/10.1121/1.3664087

McQueen, J. M., Tyler, M. D., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development*, *8*, 317–339. https://doi.org/10.1080/15475441.2011.641887

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. https://doi.org/10.1111/cdev.12087

Munson, B. (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research*, *47*, 58–69. https://doi.org/10.1044/1092-4388(2004/006

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, *21*, 124–139. https://doi.org/10.1044/1058-0360(2011/11-0009)

Nam, Y., & Polka, L. (2016). The phonetic landscape in infant consonant perception is an uneven terrain. *Cognition*, *155*, 57–66. https://doi.org/10.1016/j.cognition.2016.06.005

National Institute of Child Health & Human Development Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. National Institute of Child Health and Human Development Early Child Care Research Network. *Child Development*, *71*(4), 960–980. https://doi.org/0009-3920/2000/7104-0015

Newman, R. S. (2011). 2-year-olds' speech understanding in multitalker environments. *Infancy*, *16*(5), 447–470. https://doi.org/10.1111/j.1532-7078.2010.00062.x

Newman, R. S., Morini, G., Kozlovsky, P., & Panza, S. (2018). Foreign accent and toddlers' word learning: The effect of phonological contrast. *Language Learning and Development*, *14*, 97–112. https://doi.org/10.1080/15475441.2017.1412831

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*, 62–85. https://doi.org/10.1177/0261927X99018001005

Nishi, K., Lewis, D. E., Hoover, B. M., Choi, S., & Stelmachowicz, P. G. (2010). Children's recognition of American English consonants in noise. *The Journal of the Acoustical Society of America*, *127*, 3177–3188. https://doi.org/10.1121/1.3377080

Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, *118*, 2570–2578. https://doi.org/10.1121/1.2010407

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Perlman, M., Falenchuk, O., Fletcher, B., Mcmullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the Classroom Assessment Scoring System) in early childhood education and care settings and child outcomes. *PloS One*, *11*, 1–33. https://doi.org/10.1371/journal.pone.0167660

Perlman, M., Fletcher, B., Falenchuk, O., Brunsek, A., McMullen, E., & Shah, P. S. (2017). Child-staff ratios in early childhood education and care settings and child outcomes: A systematic review and meta-analysis. *PloS One*, *12*, 1–24. https://doi.org/10.1371/journal.pone.0170256

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, *82*, 834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Ren, J., & Morgan, J. L. (2011). Sub-segmental details in early lexical representation of consonants. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*. Retrieved from http://titan.cog.brown.edu/research/infant_lab

Rett, A., & White, K. S. (2019). Children's attention to errors in native and foreign accented speech. *Poster Presented at SRCD Biennial Meeting*. Baltimore, MD, USA.

Romeo, R., Hazan, V., & Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of the Acoustical Society of America*, *134*(5), 3781–3792. https://doi.org/10.1121/1.4824160

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, *29*, 700–710. https://doi.org/10.1177/0956797617742725

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*, 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*, 608–635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

Schafer, G., Plunkett, K., & Harris, P. L. (1999). What's in a name? Lexical knowledge drives infants' visual preferences in the absence of referential input. *Developmental Science*, *22*, 187–194. https://doi.org/https://doi.org/10.1111/1467-7687.00067

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

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

Seidl, A., Onishi, K. H., & Cristia, A. (2014). Talker variation aids young infants' phonotactic learning. *Language Learning and Development*, *10*, 297–307. https://doi.org/10.1080/15475441.2013.858575

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

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

Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska Articulation Norms project: Consonant singletons. *Journal of Speech and Hearing Research*, *36*, 533–547. https://doi.org/10.1044/jshr.3603.533

Soderstrom, M., Grauer, E., Dufault, B., & McDivitt, K. (2018). Influences of number of adults and adult:child ratios on the quantity of adult language input across childcare settings. *First Language*, *38*, 563 –581. https://doi.org/10.1177/0142723718785013

Stoel-Gammon, C., & Vogel Sosa, A. (2007). Phonological Development. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 238–256). https://doi.org/10.1002/9780470757833

Stokes, S. F., & Surendran, D. (2005). Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech, Language and Hearing Research*, *48*, 577–591. https://doi.org/1092-4388/05/4803-0577

Strömbergsson, S., Wengelin, Å., & House, D. (2014). Children's perception of their synthetically corrected speech production. *Clinical Linguistics & Phonetics*, *28*, 373–395. https://doi.org/10.3109/02699206.2013.868928

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

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

Tolchinsky, L. (2004). The nature and scope of later language development. In R. A. Berman (Ed.), *Language development across childhood and adolescence* (pp. 233–247). https://doi.org/10.1075/tilar.3.15tol

Tsuji, S., Fikkert, P., Yamane, N., & Mazuka, R. (2016). Language-general biases and language-specific experience contribute to phonological detail in toddlers' word representations. *Developmental Psychology*, *52*, 379–390. https://doi.org/10.1037/dev0000093

Uttley, L., Hillairet De Boisferon, A., Dupierrix, E., Lee, K., Quinn, P. C., Slater, A. M., & Pascalis, O. (2013). Six-month-old infants match other-race faces with a non-native language. *International Journal of Behavioral Development*, *37*(2), 84–89. https://doi.org/10.1177/0165025412467583

van der Feest, S. V. H., & Fikkert, P. (2015). Building phonological lexical representations. *Phonology*, *32*, 207–239. https://doi.org/10.1017/S0952675715000135

van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, *143*, 340–350. https://doi.org/10.1037/a0032192

van Heugten, M., Krieger, D. R., & Johnson, E. K. (2015). The developmental trajectory of toddlers' comprehension of unfamiliar regional accents. *Language Learning and Development*, *11*, 41–65. https://doi.org/10.1080/15475441.2013.879636

Wang, M. D., & Bilger, R. C. (1973). Consonant confusion in noise: A study of perceptual features. *The Journal of the Acoustical Society of America*, *54*, 1248–1266. https://doi.org/10.1121/1.427152

Weatherhead, D., & White, K. S. (2016). He says potato, she says potahto: Young infants track talker-specific accents. *Language Learning and Development*, *12*, 92–103. https://doi.org/10.1080/15475441.2015.1024835

Weatherhead, D., & White, K. S. (2018). And then I saw her race: Race-based expectations affect infants' word processing. *Cognition*, *177*, 87–97. https://doi.org/10.1016/j.cognition.2018.04.004

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*, 2020–2045. https://doi.org/10.1037/xge0000014

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384. https://doi.org/10.1111/j.1467-7687.2010.00986.x

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

Yurovsky, D., Case, S., & Frank, M. C. (2017). Preschoolers flexibly adapt to linguistic input in a noisy channel. *Psychological Science*, *28*, 132–140. https://doi.org/10.1177/0956797616668557

Appendices

Appendix A (Chapter 2)

Table A1: Familiar objects and novel labels used in Experiment 1

Block 1			Block 2		
Set	Labelled object	Distractor object	Set	Labelled object	Distractor object
1	monkey	Flower	1	pants	Bunny
	ball	*Novel*		keys	*Novel*
	murg/zurk	Shirt		*murg/zurk*	Hand
2	shoe	Bike	2	phone	Truck
	dog	*Novel*		cookie	*Novel*
	semp/neech	Balloon		*semp/neech*	Bird
3	vacuum	Banana	3	mouth	Brush
	book	*Novel*		fish	*Novel*
	gorp/tilk	Horse		*gorp/tilk*	Block
4	cat	Doll	4	cup	Hat
	foot	*Novel*		sock	*Novel*
	tibble/boogle	Spoon		*tibble/boogle*	Chair

Note. Within a set, distractor objects were counterbalanced across participants. The order of the 2 blocks, and the particular speaker (child, adult) assigned to each block, were counterbalanced across participants. Novel labels are presented as child/adult productions.

Table A2: Familiar, mispronounced, and novel labels used in Experiment 2

Set	Correct / Novel Label	Mispronounced Labels[a]		
		1-feature	**2-feature**	**3-feature**
1	monkey	Bunky	gunky	Tunky
	ball	Gall	nall	Sall
	mouth	Nouth	pouth	Kouth
	fish	pish	zish	Gish
	tibble	familiar distractor: chair		
2	shoe	foo	voo	Goo
	pants	tants	sants	Nants
	dog	nog	vog	Fog
	keys	gees	dees	Zees
	gorp	familiar distractor: doll		
3	sock	zock	pock	Bock
	book	gook	nook	Sook
	vacuum	zacuum	pacuum	Tacuum
	cup	gup	shup	Vup
	semp	familiar distractor: jacket/coat		
4	cat	tat	dat	Zat
	foot	soot	buut	Guut
	phone	pone	tone	Doan
	cookie	pookie	dookie	Mookie
	murg	familiar distractor: block		

Note. Toddlers heard each familiar label produced with one of 4 pronunciations (correct or with a 1-, 2-, or 3-feature onset mispronunciation). Within a set, the particular items assigned to each pronunciation were counterbalanced across participants.

[a.] rimes are the same as those of the correct productions regardless of spelling

Supplemental Material: Production details for the child and adult speakers (Experiment 1)

In addition to general acoustic measures, we also looked at the child vs. adult productions of some specific representative sounds, to get a sense of how on-target the child's productions were. Because the stimulus set was not designed specifically with these analyses in mind (and was designed instead based on word familiarity), the words began with a range of onset consonants. This, unfortunately, meant that we did not have enough tokens of specific sounds to conduct any acoustic analyses. Nonetheless, Table S1 provides values for four stop and fricative onset consonants. Given that the child consistently used a faster speaking rate than the adult, data are present both as the actual duration and as C/V ratios that take speaking rate into account. There are both similarities and differences between the adult and child productions. Like the adult speaker, the child speaker made a clear distinction between her voiced and voiceless stops. However, despite similarities in the accuracy (center of gravity) of her /s/ and /ʃ/ productions, the duration of her /ʃ/ was considerably longer than the adult speaker's, while the duration of her /s/ was somewhat shorter.

Table S1: Acoustic values for sample productions from child and adult speakers

	Child Speaker			Adult Speaker		
Phoneme	**Duration (ms)**	**C/V ratio[a]**	**Centre of Gravity (Hz)**	**Duration (ms)**	**C/V ratio**	**Centre of Gravity (Hz)**
Stops (VOT)						
/b/ in book	5.16	0.04	N/A	15.15	0.09	N/A
/k/ in keys	134.6	0.39	N/A	124.12	0.39	N/A
Fricatives						
/ʃ / in shoe	228.77	0.89	3,996.1	204.12	0.53	3,772.5
/s/ in sock	210.93	0.90	5,939.6	240.65	1.19	5,418.7

[a] C/V ratio is a measure of duration that accounts for differences in speaking rate

Supplemental Material: Timecourse Plots

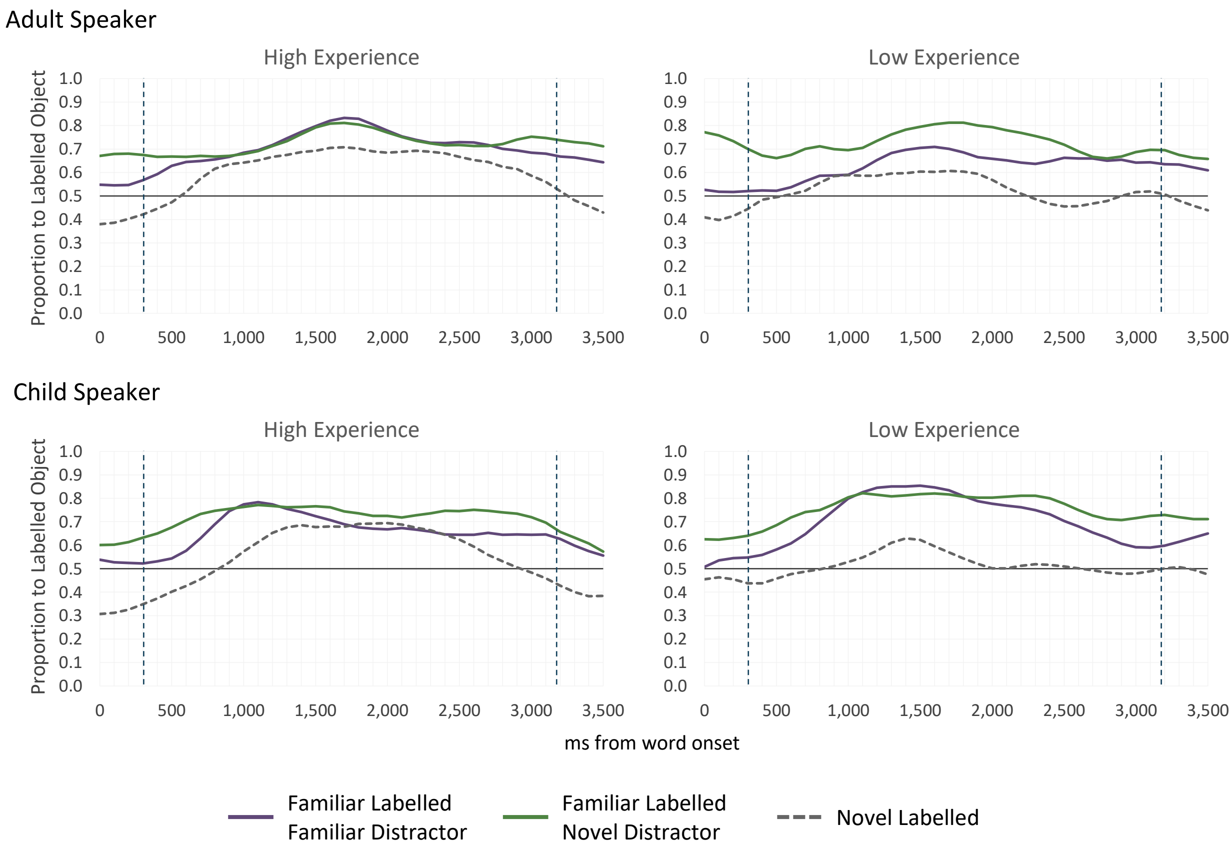


Figure S1: Timecourse plots for Experiment 1 highlighting the difference in novel label processing for High vs. Low Experience groups, and the similarity in familiar label processing for Child and Adult Speakers. The Y-axis represents proportion looking to the labelled object during the naming phase; the X-axis represents milliseconds from word onset.

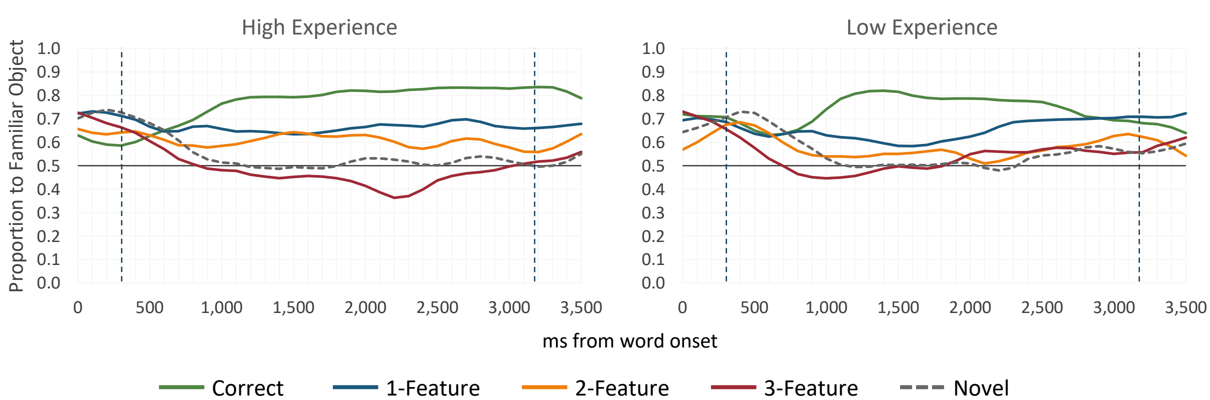


Figure S2: Timecourse plots for Experiment 2, demonstrating the sensitivity of both Experience groups to mispronunciations, but also highlighting the greater sensitivity exhibited by the High Experience group. The Y-axis represents proportion looking to the familiar object during the naming phase; the X-axis represents milliseconds from word onset.

Appendix B (Chapter 3)

Table B1: Object labels and pronunciations

Contrast	Labelled Object	Mispronunciation Condition	
		Common	Infrequent
Familiar Object Labels			
/t/-/ʧ/	table	CP: [tebl]	MP: [ʧebl]
	turtle	CP: [tɝtl]	MP: [ʧɝtl]
	chair	MP: [tɛɹ]	CP: [ʧɛɹ]
	cheese	MP: [tiz]	CP: [ʧiz]
/s/-/ʃ/	sock	CP: [sɑk]	MP: [ʃɑk]
	soap	CP: [soʊp]	MP: [ʃoʊp]
	shoe	MP: [su]	CP: [ʃu]
	shirt	MP: [sɝt]	CP: [ʃɝt]
/w/-/ɹ/	water	CP: [wɑtɚ]	MP: [ɹɑtɚ]
	wagon	CP: [wægʌn]	MP: [ɹægən]
	raisin	MP: [wezɪn]	CP: [ɹezɪn]
	rock	MP: [wɑk]	CP: [ɹɑk]
/w/-/l/	wheel	CP: [wil]	MP: [lil]
	window	CP: [wɪndoʊ]	MP: [lɪndoʊ]
	lamp	MP: [wæmp]	CP: [læmp]
	lion	MP: [waɪʌn]	CP: [laɪʌn]
Novel Object Labels [a]			
---	gimble	[gɪmbl]	bird
	finkle	[fɪŋkl]	cookie
	poggle	[pɑgl]	cat
	moogie	[mugi]	book
Filler Trials [a]			
---	cup	[kʌp]	horse
	ball	[bɑl]	car
	bike	[baɪk]	hat
	keys (toy)	[kiz]	phone (toy)

Note. MP = mispronunciation; CP = correctly pronounced

[a] familiar distractors for novel label and filler trials are provided to the right of each labelled object

Appendix C (Chapter 4)

Table C1. Scripts used by the adult and child speakers that include the child's /ɹ/→[w] mispronunciation of familiar object labels

	Adult Speaker	Child Speaker
Training		
trial 1	Look! A roogie. Do you see the roogie? That's right. It's a roogie. Roogie.	---
trial 2	Wow! A roogie. Do you see the roogie? Look! It's a roogie. Roogie	---
Familiarization		
rainbow	Cool! A rainbow. Look! It's a rainbow.	Cool! A wainbow. Do you see the wainbow?
Rooster	Look! A rooster. Do you see the rooster?	Ooo! A wooster. Do you see the wooster?
Rattle	Wow! A rattle. Do you see the rattle?	Wow! A wattle. Do you see the wattle?
Rabbit	Look! A rabbit. Wow! A rabbit.	Cool! A wabbit. Do you see the wabbit?
Test		
Book	---	Do you see the book?
Woogie	Ooo! A woogie.	Ooo! A woogie

Table C2. Acoustics of the audio for the adult speaker who produced familiar object labels correctly, and the child speaker who mispronounced familiar object labels

	Adult Speaker			Child Speaker		
	duration (ms)	**Intensity (dB SPL)**	**mean f0 (Hz)**	**duration (ms)**	**Intensity (dB SPL)**	**mean f0 (Hz)**
Training						
trial 1	7314	52.6	201	---	---	---
trial 2	7460	52.9	205	---	---	---
Familiarization						
rainbow	4158	53.1	197	3450	53.5	236
Rooster	3525	52.7	205	2746	53.7	221
Rattle	3797	52.7	208	2891	53.6	241
Rabbit	3929	52.9	215	3565	53.3	239
Test						
Book	---	---	---	1520	58.7	237
Woogie	1727	56.2	193	1598	57.8	186

Note. training and familiarization videos were 5.5 seconds longer than the audio (1 s before, 4.5 s after)