ABSTRACT

Title of Dissertation: ISOLATING EFFECTS OF PERCEPTUAL ANALYSIS AND SOCIOCULTURAL CONTEXT ON CHILDREN'S COMPREHENSION OF TWO DIALECTS OF ENGLISH, AFRICAN AMERICAN ENGLISH AND GENERAL AMERICAN ENGLISH

Michelle Elizabeth Erskine, Doctor of Philosophy, 2023

Dissertation directed by: Professor Jan. R. Edwards
Associate Professor Yi Ting Huang,
Department of Hearing and Speech Sciences

There is a long-standing gap in literacy achievement between African American and European American students (e.g., NAEP, 2019, 2022). A large body of research has examined different factors that continue to reinforce performance differences across students. One variable that has been a long-term interest to sociolinguists and applied scientists is children's use of different dialects in the classroom. Many African American students speak African American English (AAE), a rule-governed, but socially stigmatized, dialect of English that differs in phonology, morphosyntax, and pragmatics from General American English (GAE), the dialect of classroom instruction. Empirical research on dialect variation and literacy achievement has demonstrated that linguistic differences between dialects make it more difficult to learn to read (Buhler et al., 2018; Charity et al., 2004; Gatlin & Wanzek, 2015; Washington et al., 2018, *inter alia*) and recently, more difficult to comprehend spoken language (Byrd et al., 2022, Edwards et

al., 2014; Erskine, 2022a; Johnson, 2005; de Villiers & Johnson, 2007; JM Terry, Hendrick, Evangelou, et al., 2010; JM Terry, Thomas, Jackson, et al., 2022).

The prevailing explanation for these results has been the ***perceptual analysis hypothesis***, a framework that asserts that linguistic differences across dialects creates challenges in mapping variable speech signals to listeners' stored mental representations (Adank et al., 2009; Clopper, 2012; Clopper & Bradlow, 2008; Cristia et al., 2012). However, spoken language comprehension is more than perceptual analysis, requiring the integration of perceptual information with communicative intent and sociocultural information (***speaker identity***). To this end, it is proposed that the perceptual analysis hypothesis views dialect variation as another form of signal degradation. Simplifying dialect variation to a signal-mapping problem potentially limits our understanding of the contribution of dialect variation to spoken language comprehension. This dissertation proposes that research on spoken language comprehension should integrate frameworks that are more sensitive to the contributions of the sociocultural aspects of dialect variation, such as the role of linguistic and nonlinguistic cues that are associated with speakers of different dialects.

This dissertation includes four experiments that use the visual world paradigm to explore the effects of dialect variation on spoken language comprehension among children between the ages of 3;0 to 11;11 years old (years;months) from two linguistic communities, European American speakers of GAE and African American speakers with varying degrees of exposure to AAE and GAE. Chapter 2 (Erskine [2022a]) investigates the effects of dialect variation in auditory-only contexts in two spoken word recognition tasks that vary in linguistic complexity: a) word recognition in simple phrases and b) word recognition in sentences that vary in semantic predictability. Chapter 3 [Erskine (2022b)] examine the effects of visual and auditory speaker identity cues on dialect variation on literal semantic comprehension (i.e., word recognition in semantically facilitating sentences). Lastly, Chapter 4 [Erskine (2022c)] examines the effects of visual and auditory speaker identity cues on children's comprehension of different dialects in a task that evaluates pragmatic inferencing (i.e., scalar implicature). Each of the studies investigate the validity of the perceptual analysis against sociolinguistcally informed hypotheses that account for the integration of linguistic and nonlinguistic speaker identity cues as adequate explanations for relationships that are observed between dialect variation and spoken language comprehension.

Collectively, these studies address the question of how dialect variation impacts spoken language comprehension. This dissertation provides evidence that traditional explanations that focus on perceptual costs are limited in their ability to account for correlations typically reported between spoken language comprehension and dialect use. Additionally, it shows that school-age children rapidly integrate linguistic and nonlinguistic socioindexical cues in ways that meaningfully guide their comprehension of different speakers. The implication of these findings and future research directions are also addressed within.

ISOLATING EFFECTS OF PERCEPTUAL ANALYSIS AND SOCIOCULTURAL CONTEXT ON CHILDREN'S COMPREHENSION OF TWO DIALECTS OF ENGLISH, AFRICAN AMERICAN ENGLISH AND GENERAL AMERICAN ENGLISH

by

Michelle Elizabeth Erskine

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Dr. Jan Edwards, Chair
Dr. Yi Ting Huang, Co-Chair
Dr. Matthew Goupell
Dr. Rochelle Newman
Dr. Jared Novick
Dr. Shenika Hankerson, Dean's Representative

Dedication

This dissertation is dedicated to Yeshua, my heavenly and spiritual father from whom all blessings flow. Thank you for walking with me through each phase of this journey and for leading me towards a community of family, friends, and colleagues who have supported, uplifted, and guided me each step of the way.

"In all your ways acknowledge Him and He shall direct your paths."

Proverbs 3:6

Acknowledgements

This has been one of the most difficult sections to write because of my own anxieties about forgetting to acknowledge the many individuals who generously provided their knowledge, expertise, and unwavering support. There are not enough words to express my sincere gratitude and appreciation that I have for each person who has taken this journey with me including those that have been present since the onset of my academic pursuits and those who have joined me at various points along the way. You have all been a voice of reason, encouragement and guidance.

There are so many individuals that I want to especially thank. First, I would like to thank my primary advisor Professor Jan Edwards for her invaluable words of advice and inspiration, and the generous help you provided me over the course of 8 years of working together. I could not have asked for a better mentor, teacher and friend. This journey has at times felt like climbing a ladder with several missing rungs, but you were there for each step providing the necessary scaffolding while guiding me towards independence. Our conversations led to many insightful realizations that have transformed my outlook and appreciation for language variation research, as well as my steadily growing appreciation for *high-frequency* words. I am hugely indebted to your generosity and patience in helping me become a scientist that is better equipped to address questions related to educational equity, dialect variation, and the influence of social properties of speakers on academic achievement.

I would also like to thank my co-advisor Associate Professor Yi Ting Huang. From the earliest points of my doctoral journey at the University of Maryland, you helped to further develop my scientific skills and guided me towards addressing important gaps in the dialect literature. Furthermore, you encouraged me to break away from deficit driven narratives that are often used to characterize and explain relationships between dialect variation and academic achievement. Thank you for helping me to the push boundaries of my own conceptualizations of the literature and for encouraging me to pursue cutting edge questions at the intersection of dialect variation and the semantic-pragmatic interface. Without your supervision, I would not have pushed myself beyond the realm of literal semantic comprehension.

I also want to express my thanks to Professor Rochelle Newman, Professor Matthew Goupell, Associate Professor Jared Novick, and Assistant Professor Shenika Hankerson for reading my thesis and providing enlightening and insightful comments. I especially want to thank Professor Rochelle Newman for the provision of additional resources and support which aided data collection during an extremely challenging period of our history. This work would not have been possible without your generous extension of this lifeline.

In addition, I would like to thank members of the Learning to Talk Lab and colleagues in my cohort for their willingness to listen to my ideas and provide extremely helpful feedback. I also want to thank those who aided in data collection and/or analysis. A massive shout-out and thanks is owed to the entire team including: Tatiana Thonesavanh, Zachary Maher, Christina Blomquist, Arynn Byrd, Kathleen Oppenheimer, Allison Johnson, Meg Cychosz., Robyn Toler, Megan Park, Victoria Dubin, Astrid Widjaja, and Kimberlie Phung Tran.

On a personal level, my deepest and most sincere appreciation is reserved for my parents Faith and K-D Erskine for their unconditional love, support, and confidence in me during the course of completing this PhD. Words cannot begin to articulate how much of an inspiration you are to me. I am in awe of the sacrifices each of you have made in leaving your home country for a land filled with uncertainty, racial oppression, and a very sparse community of friends or family. You are both my strength and foundation. I could not have even imagined this life without your endless words of encouragement (and the occasional bouts of Caribbean-based tough love). Thank you for pushing me to realize that I can overcome any adversity that life throws my way and for showing me that it is faith (mostly crazy faith) that paves the way ahead. I am so blessed to have you as my parents. I also send thanks to my *chidera,* my one and only sister, Melissa. You provided endless amounts of support from the start until the very end of this academic journey. I will never forget your willingness to listen to my research ideas, practice talks, and remain on the phone with me as I took very long meditative walks. Finally, thanks to my friends both here and overseas. You were all so understanding when I missed important social events, forgot to return several phone calls or respond to texts, or when our conversations became mostly about my research or some article I read earlier that week. Each of you mean so much to me and were a crucial part of helping me maintain my sanity.

This work was supported in part by:

NIDCD R01 02932 grant to Jan Edwards, Mary E. Beckman, and Benjamin Munson;

NICHD P30 HD03352 grant to the Waisman Center;

T32 Training Grant DC -5359-10 to Susan Ellis Weismer;

NSF grant #1449815 to Colin Phillips;

NICHD F31HD1079730-02 to Michelle Erskine.

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1: Introduction

There is a long-standing gap in academic achievement between African American and European American students, a gap that has recently widened because of the pandemic and protracted periods of remote schooling (NAEP, 2022). Among several social and economic factors that likely contribute to its persistence, sociolinguists and educators have proposed that dialect variation partially explains the gap's persistence. This is because many African American students speak African American English (AAE), a rule governed, but highly stigmatized, dialect of English that differs in phonology, morphosyntax, and pragmatics from the dialect of classroom instruction, General American English (GAE).

Decades of empirical research on dialect variation and literacy achievement have demonstrated that linguistic differences between AAE and GAE make it more difficult to learn to read (Buhler et al., 2018; Charity et al., 2004; Gatlin &Wanzek, 2015; Washington et al., 2018, *inter alia*) and recently, more difficult to comprehend spoken language (Byrd et al. 2022, Edwards et al., 2014; Erskine, 2022a; Johnson, 2005; de Villiers et al., 2007; JM Terry, Hendrick, Evangelou, et al., 2010; JM Terry, Thomas, Jackson, et al., 2022). This shift in attention is important because the majority of children's early learning experiences occur through the medium of spoken language. For example, in classroom contexts teachers often provide their students with instructions and engage in critical thinking opportunities through dialogue. Moreover, early academic tasks depend a great deal on children having robust oral language skills including spoken language comprehension (Lervåg et al., 2018; Metsala et al., 2021; Wise et al., 2007). A fundamental aspect of learning depends on children's ability to efficiently and accurately encode information from spoken input.

The majority of the research on spoken language comprehension and dialect variation is correlational, and clear causal explanations have yet to be identified. Nevertheless, the prevailing explanation for the influence of dialect variation on spoken language comprehension has been the ***perceptual analysis hypothesis*** (Adank et al. 2009; Clopper, 2012; Clopper & Bradlow, 2008; Cristia et al. 2012). This framework asserts that linguistic differences between the dialects creates another layer of difficulty in mapping an unfolding speech signal to listeners' existing mental representations. The perceptual analysis hypothesis operates as a variant of the linguistic interference hypothesis but instead of focusing on speech-to-print mismatches, it targets the mismatches between the auditory signal and listeners' stored mental representations. Furthermore, the perceptual analysis hypothesis treats dialect variation as a form of signal degradation in which discrepancies between the dialects result in poorer perceptual analysis, negatively impacting the accuracy and efficiency of spoken language comprehension (Adank & McQueen, 2007; Adank et al., 2009; Clopper, 2012; Clopper & Bradlow, 2008; Erickson et al. 2017; Lanwermeyer et al., 2016; Mattys et al. 2012; McMillan et al., 2016).

However, the critiques described above also hold true for the perceptual analysis hypothesis. It too neglects the fact that spoken language comprehension is more than analysis of the speech signal. While engaging in rapid perceptual analysis, listeners simultaneously attend to and integrate higher-level contextual information about speaker identity and the speaker's communicative intent. Moreover, dialect variation is inherently social and is usually nested in communities of speakers that vary systematically by different social factors. Additionally, previous research has shown that listeners, adults and children, are sensitive to linguistic and non-linguistic properties of speakers and use their knowledge and experience with different speakers to facilitate spoken language comprehension (Babel, 2022; Babel & Russell, 2015;

Kang & Rubin, 2009; Staum Casasanto, 2008). Despite the evidence that listeners integrate speaker identity in ways that meaningfully guide spoken language comprehension, very few studies on children's comprehension of dialect differences have taken into account the social nuances of dialect variation.

Integration of speaker identity and communicative intent for the purpose of spoken language comprehension is not considered by the perceptual analysis hypothesis. The results of the studies in this thesis suggest that a valid theory of real-time spoken language comprehension in the face of dialect variation must account for how listeners avail themselves to both linguistic and non-linguistic cues about the speaker to guide their comprehension. Moreover, it proposes that comprehending speakers of different dialects depend on unique interactions between the linguistic signal and additional socioindexical properties of the speaker (e.g., age, gender, race/ethnicity) (Hay et al., 2006; Johnson et al., 1999; Shafto et al., 2012; Weatherhead & White, 2018; Xiao et al., 2018).

The primary objective of this thesis was to consider the role of perceptual analysis in contexts that differ in the availability of additional non-linguistic socioindexical cues about the speaker's identity. A dearth of research has been conducted in service of empirically evaluating the validity of either of these hypothesis as a causal explanation for relationships between dialect use and spoken language processing. Using the visual world paradigm, a measure that is sensitive to the real-time demands of spoken language processing (i.e., the relative ease of word recognition and/or the consideration of possible interpretations), four experiments were used to examine the contribution of dialect variation to spoken language comprehension in two groups of children. One group was European American children who were primarily exposed to GAE, while the other group was African American children who had varying degrees of exposure to both AAE and GAE. These experiments included stimuli that varied in linguistic complexity and access to cues about the speaker's identity. The role of dialect variation was examined in three linguistic contexts: a) word recognition in simple phrases, b) word recognition in longer utterances that contained semantically facilitating or neutral verbs, and c) sentences with implicit, pragmatic interpretations. Additionally, this dissertation compared children's sensitivity to dialect variation in two types of visual world paradigm experiments across three chapters.

Chapter 2 examines the influence of dialect variation in the context of auditory-only cues to speaker identity across two experiments: word recognition in simple phrases and sentences that vary in semantic predictability (Erskine, 2022a). Chapter 3 examines the influence of dialect variation in the presence of visual and auditory cues to speaker identity during word recognition of semantically facilitating sentences (Erskine, 2022b). Chapter 4 (Erskine, 2022c) builds on the results of chapter 3 by examining the effects of speaker identity in nonliteral sentence contexts which require greater integration of speaker identity cues than literal-semantic tasks which were the focus of chapters 2 and 3. Lastly, chapter 5 provides an interwoven discussion of the results across all three chapters providing further evidence that it is theoretically flawed to presume that dialect variation is another form of signal degradation. Additional implications for future research and paths forward for understanding effects of dialect variation are also considered in chapter 5.

Chapter 2: Examining the influence of dialect variation in spoken word recognition and semantic prediction

Experiment 1

There is a long-standing gap in literacy achievement between African American and European American students (e.g., NAEP, 2019, 2022). A large body of research has examined different factors that continue to reinforce performance differences across students. One variable that is of overwhelming interest to sociolinguists and other applied scientists is children's use of different dialects in the classroom. Many African American students speak African American English (AAE), a rule-governed, but socially stigmatized, dialect of English that differs in phonology, morphosyntax, and pragmatics from General American English (GAE), the dialect of classroom instruction. An abundance of research beginning in the early 1970s has shown consistent correlational evidence that linguistic differences between AAE and GAE are related to academic achievement among African American students (Charity et al. 2004; Connor & Craig, 2006; Craig & Washington, 2004; Gatlin & Wanzek, 2015, Goodman & Buck, 1973; Harber, 1977; Hart et al. 1980; JM Terry, Hendrick, Evangelou, et al., 2010, Melmed, 1973; NP Terry et al., 2012; NP Terry et al., 2016; Seymour & Ralabate, 1985; Simons & Johnson, 1974). Across these studies, it has been consistently observed that children who produce higher rates of AAE have lower math, language, and literacy scores (e.g., Gatlin & Wanzek, 2015; JM Terry, Hendrick, Evangelou, et al., 2010; Washington et al. 2018). The vast majority of this research has focused on establishing relationships between dialect use and learning to read (i.e., phonological awareness, decoding, spelling, and more recently, reading comprehension), largely because of concerns about achievement differences in reading between African American and European American children. Such studies have shown that dialect differences affect decoding and spelling by creating less transparent mappings between the phonological and orthographic representations of words (Kohler et al. 2007; Thomas-Tate et al., 2004; Treiman & Kessler, 2004). More recent studies have shown that dialect use even impacts higher-level reading processes such as reading comprehension. Specifically, through concurrent and longitudinal research, Terry and colleagues (NP Terry, Connor, Petscher, et al., 2012; NP Terry, Connor, Johnson, et al., 2016) demonstrated that children's continued use of high rates of AAE across grades (i.e., less dialect-shifting) results in lower reading comprehension scores. Terry and colleagues suggest that mismatches which have been shown to primarily impact speech-to-print mapping for decoding and spelling also have downstream consequences for the comprehension of text. Jointly, these studies demonstrate that dialect use and/or dialect-shifting plays an important role in understanding reading achievement among student speakers of nonmainstream varieties of English, like AAE. Moreover, this correlational evidence seems particularly convincing as these effects are often observed even when the studies control for children's linguistic/literacy knowledge and socioeconomic status (Gatlin & Wanzek, 2015; Terry et al. 2016; Washington et al. 2018).

However, continued research on reading development only partially address the scale of influence of dialect use on academic achievement. Moreover, the majority of evidence in this area of research is primarily correlational, meaning that explanatory mechanisms linking dialect use to academic achievement remain unclear, even for reading. Therefore, it is important to consider other variables related to achievement that are potentially impacted by dialect variation. One factor of interest that is slowly gaining momentum in the literature is the effect of dialect use on ***spoken language comprehension***.

From a practical viewpoint, examining spoken language comprehension provides ways for researchers to capture relevant sources of variation in academic performance that would otherwise be obscured if research focused only on measures of reading. From the earliest moments of children's educational experiences, learning occurs through the medium of spoken language. For example, in classroom contexts teachers often provide their students with instructions and engage in critical thinking opportunities through dialogue. Moreover, early academic tasks depend a great deal on children having robust oral language skills including spoken language comprehension (Lervåg et al., 2018; Metsala et al., 2021; Speece et al., 1999; Wise et al., 2007). Therefore, a foundational aspect of learning, or at the very least, accessing the curriculum content, relies on students' ability to efficiently and accurately encode information from spoken input.

From a theoretical perspective, existing explanations for the relationship between dialect use and linguistic comprehension of GAE posit that students bear an additional cognitive load in resolving speech-to-print or speech-to-speech mismatches, which may adversely affect children's performance on academic measures (Edwards et al., 2014; JM Terry, Hendrick, Evangelou, et al., 2010, JM Terry, Thomas, Jackson, et al., 2022). For example, one explanation that has been widely acknowledged is the ***linguistic mismatch (or interference) hypothesis***. This account asserts that AAE listeners are more likely to experience difficulty learning to read because dialect differences between AAE and GAE create lexical confusability (Hartwell, 1980; Piestrup, 1973; Troutman & Falk, 1982; Simons & Johnson, 1974; Wolfram, 1971). Another explanation, not necessarily orthogonal to the interference hypothesis, is the ***linguistic flexibility hypothesis*** which suggests that reading is impacted not solely by children's use of AAE but rather their flexible use and awareness of GAE (Connor, 2008; Craig et al., 2014; Holt et al. 2021; Mitri & NP Terry, 2014; NP Terry, 2008). In other words, children with greater linguistic knowledge of GAE are less likely to experience interference from dialect differences between AAE and GAE, and are thus more likely to have higher reading scores (Charity et al. 2004; Gatlin & Wanzek, 2015). Undergirding both of these explanations is an implicit assumption that linguistic mismatch creates an additional challenge, one that may be lessened through increased awareness of GAE, but nonetheless remains present as AAE-speaking children learn to switch between their native dialect and varieties that are closely aligned with the language of instruction (i.e., GAE).

Empirically, some studies provide indirect evidence for a relationship between dialect use and spoken language comprehension in school-age children. Edwards et al. (2014) examined the comprehension of words that were lexically ambiguous or unambiguous among four- to eight-year-old AAE-speaking children. For example, words like *gold* would be considered ambiguous as they could be perceived as either *goal* [goʊl] or *gold* [goʊld] among AAE speakers. Edwards and colleagues found that AAE-speaking children were overall less accurate in comprehending words that were lexically ambiguous relative to unambiguous words that did not differ between AAE and GAE (e.g., *doll*). Earlier studies on the comprehension of third person singular -s show similar findings. For example, Johnson (2005) and de Villiers and Johnson (2007) both tested children ranging from age 3 to 7 on their comprehension of tense/aspect and number agreement that is cued by third person singular -s. Across these studies they observed that AAE-speaking children did not comprehend this feature, while GAE-speakers began to show sensitivity to its grammatical interpretations approximately at age 5 to 6. These findings were later corroborated by Beyer and Hudson-Kam (2012) who assessed first and second grade AAE-speaking children on their comprehension of a variety of GAE morphological markers. They found that while GAE-speaking children demonstrated comprehension of some tense markers (i.e., past tense -ed, future contractable ‘ll, and third person -s), AAE-speaking children only showed accurate comprehension of morphological forms that overlapped between GAE and AAE (i.e., plural -s).

Altogether, these studies provide preliminary evidence that linguistic differences across dialects do have the potential to influence comprehension. However, the mechanism(s) underlying this connection are still poorly understood.

The results of the studies mentioned above are potentially consistent with a perceptual analysis framework of spoken language processing. Psycholinguistic models of spoken language processing indicate that accurate and efficient language comprehension depends on the listeners' ability to create associations between a rapidly unfolding signal and underlying phonological and lexical representations. **The *perceptual analysis hypothesis*** posits that linguistic processing unfolds with relative ease when interlocutors speak the same dialect or language variety, but this process will be disrupted when there is an asymmetry between the dialects or language varieties spoken (e.g., as is the case between AAE and GAE) (Cristia et al., 2012). The perceptual analysis hypothesis is grounded in the literature on spoken language comprehension more generally and so it takes into account the linguistic and experiential factors that might influence the effect of dialect variation on linguistic processing. For instance, this hypothesis predicts that linguistic processing will be disrupted in the presence of linguistic differences. Since dialect variation impacts many levels of language including prosody, phonology, and morphosyntax, linguistic differences at any of these levels may make processing more difficult. Furthermore, this hypothesis also anticipates that the effects of dialect variation will be context-specific, occurring in some language environments but not others. That is, effects of dialect variation on perceptual processing may be more pronounced as linguistic demands increase, such as when there is ambiguity among candidate lexical items (by increasing lexical competition among words), or when morphosyntax is used differently between the two dialects in question, or when the environment is particularly noisy. Furthermore, the impact of dialect differences on perceptual processing may be more evident when listeners have less linguistic experience with one of the dialects in question. The predictions proffered by the perceptual analysis hypothesis take into account both the linguistic context as well as the linguistic experience of the listener.

A perceptual analysis explanation differs from previous hypotheses in accounting for effects of dialect difference on spoken language processing (i.e., linguistic flexibility) because it highlights a specific linking explanation making clearer predictions about why and where differences in spoken language processing will arise (e.g., encoding of linguistic information or retrieval of relevant linguistic meanings). Intuitively, we can see how this framework may be suitable for evaluating and potentially explaining, the relationship between dialect variation and spoken language processing. If linguistic differences disrupt perceptual analysis, then this could place greater demands on cognitive resources typically engaged during later stages of comprehension or prevent children from successfully completed academic-related tasks. Additionally, the contexts in which dialect differences are observed may vary according to the salience of the linguistic differences (e.g., when dialect differences are more prominent or lexical ambiguities are more likely to occur). This type of explanation aligns with the speculation from the sociolinguistic and education research that greater processing costs are incurred when processing an unfamiliar dialect, but has not been empirically tested.

To date, there have been no studies directly investigating the validity of a perceptual analysis account as a valid explanation for any relationships between spoken language processing of AAE and GAE among children. However, there has been some research on accent variation to support a relationship between linguistic differences and processing. Much of this work has focused on either infants or adults, but there is a small body of research on school-age children and accent variation processing. Using the visual world paradigm, Creel (2012) examined three- to five-year-old's recognition of familiar words in three accent conditions: a) canonical ("fish", [fɪʃ]), up-shifted vowel ("feesh" [fiʃ]), and down-shifted vowel ("fesh" [fɛʃ]) productions of words. The results showed that while children accurately identified words across the different "accent" conditions, children were slower to look to the target image in up- and down-shifted accent conditions relative to canonical word productions, which would be most familiar to children. Further, two additional studies that focused on comprehension accuracy also found that accent differences influenced spoken language comprehension.

Barker and Meyer-Turner (2015) evaluated the effect of listening to a non-native accent on two experimental tasks that differed in linguistic complexity: word recognition and story comprehension. In the word recognition task, children heard isolated words and were asked to select one of four images corresponding to the target word heard. In the sentence comprehension task, children were asked to select a picture that matched a question related to story prompt heard. Children between 3- and 6-years-old heard words in either a familiar accent (American English) or in a non-native accent (Indian English). The results showed that children's word recognition comprehension accuracy was lower in the non-native accent condition. Story comprehension also showed an effect of accent familiarity but in the opposite direction, in which performance was higher in the non-native accent condition. The authors explain that the difference in task demands and in the prosodic features of the words within each passage resulted in performance differences. In the word recognition task, children did not have time to acquaint themselves with the non-native accent. However, in story comprehension listeners may have had sufficient exposure to the accent differences which may have facilitated their comprehension. Additionally, there may have been differences in speech rate across the accent conditions; the passages in Indian English were produced with a slower speaking rate and prosodic differences which may have increased children's attention to the stories, resulting in more accurate responses to the comprehension questions.

Nathan, Wells, and Donlan (1998) used a definition task to assess children's access to lexical representations across two dialects of English with highly salient linguistic differences: Glaswegian English (unfamiliar) and British English (familiar). Using a between-subjects design, 4- and 7-year-old British English-speaking children heard words in either the familiar or unfamiliar accent. Nathan et al. hypothesized words in the non-native accent would result in greater lexical ambiguity making it harder for children to retrieve the correct definition for a word, leading to a greater number of errors in the unfamiliar accent condition. They found that children had poorer comprehension of words in the unfamiliar accent than in their native accent. Moreover, older children produced fewer errors overall than younger children, suggesting a developmental timescale for accent variation processing, which is partially attributed to a maturation of children's linguistic and cognitive systems.

Though few studies have evaluated how linguistic differences between AAE and GAE impact spoken language comprehension, the research on language variation (accent and regional dialect variation) provides some clear predictions. First, these studies suggest that when linguistic differences are prominent and listeners have less experience with and/or exposure to dialect/accent variation, linguistic differences will create uncertainty in recognizing words during spoken language processing. This perceptual uncertainty may at times be realized as poorer accuracy on tasks such as word recognition or defining lexical terms. However, it may also be realized as a delay in the time course of real-time processing as was shown by Creel (2012). Second, as children’s linguistic and cognitive skills mature, their ability to process and extract meaning from unfamiliar/non-native accents improves. Therefore, it is plausible that growth in vocabulary, cognitive, and other metalinguistic skills help children to more quickly associate dialect variants as another form for a known word in their lexicon (Nathan et al., 1998; Fernald et al., 2008). For example, some children showed increased adaptation after increased periods of exposure to variation (Barker & Turner, 2015) and older children demonstrated improved accuracy in a less familiar accent (Nathan et al., 1998). These findings also suggest that at least some linguistic differences between dialects on their own might not disrupt processing. However, specific contexts, such as those that highlight dialect use to a greater degree, may be important for observing effects of dialect variation on spoken language processing, a point we return to later in the discussion section.

Using a perceptual analysis framework, this paper sought to understand how differences between a stigmatized dialect of English (AAE) and a dominant or widely accepted dialect of English (GAE) are processed by children belonging to two different linguistic communities, children who speak AAE at home and children who speak GAE. Previous research has demonstrated that variation due to sociolects or accents influence children's spoken word recognition (Edwards et al. 2014; Barker & Turner, 2015; Creel, 2012). This paper extends the existing research by using the visual world paradigm to assess real-time spoken word recognition under two dialect conditions, hearing AAE and hearing GAE. We chose this paradigm because it is sensitive to both the accuracy and time course of word recognition.

As discussed above, the perceptual analysis hypothesis predicts that linguistic differences between AAE and GAE, even differences in prosody, phonology, or acoustics can lead to processing uncertainty. In this case, the perceptual analysis hypothesis generates the following prediction: Children across different linguistic communities (i.e., speakers of GAE and speakers of AAE) will exhibit slower and less accurate comprehension of the dialect that is less familiar to them. Specifically, GAE-speaking children will demonstrate poorer word recognition in AAE and AAE-speaking children will demonstrate poorer word recognition of words in GAE because there are phonological differences between the dialects. This study is among the first to directly examine whether dialect differences affect the efficiency of real-time spoken word recognition. Two research questions were addressed. First, do linguistic differences between AAE and GAE influence spoken word recognition in two speaker groups of children ages 3-to-5 years old? Second, do individual differences in vocabulary and/or maternal education explain children's recognition of words that differ across two dialects of English, AAE and GAE?

Methods

Participants

Listeners

The participants were 56 preschool children between the ages of 3;0 and 5;11. Twenty-one of these participants were AAE-speaking children (mean age = 48 months; 14 female) and 35 were GAE-speaking children (mean age = 51 months; 26 female) recruited from community centers, preschools, and early education programs (e.g., Head Start centers) in Madison, Wisconsin. Henceforth, we will refer to all participants as *listeners* to demarcate the child participants in this study from speakers who produced the auditory stimuli in this study.

Children across linguistic communities (i.e., AAE and GAE listeners) were not evenly distributed across maternal education levels. Most AAE listeners were from families with lower levels of maternal education, while most GAE listeners were from families with higher levels of maternal education. This distribution of participants across maternal education is not unusual, instead it reflects the typical distribution of speakers of minoritized dialects across different socioeconomic levels, at least in the United States (US Census Bureau 2020). Two additional children were tested but were excluded from the analyses due to high levels of missing data (i.e., greater than 50% missing data).

Dialect group assignment among listeners

Each child's dialect was confirmed through a series of informal conversations and play-based interactions among children, caregivers, and either a research assistant who was a native speaker of AAE or a trained speech-language pathologist who had several years of experience listening to different regional varieties including AAE and GAE. According to caregiver report, all children in this sample were typically developing and had no previous history of receiving speech, language, and/or hearing services.

Maternal education among listeners

Caregivers completed a questionnaire in which they reported their family's maternal education level, which was used as a measure of socioeconomic status. All caregivers were provided with six multiple choice options about the highest level of education completed. Their responses were grouped into three categories: high (graduate degree, college degree), mid (some college/trade school/associate degree), and low (high school diploma, GED, or less). Table 1 provides information about the distribution of participants across these three categories.

Thirty of the children reported in the analyses below were part of a larger longitudinal study examining the relationships between linguistic input, vocabulary size, and auditory word recognition. Other measurements were collected as part of that longitudinal project, but were not analyzed in this paper. The analysis is restricted to the variables of interest for addressing the research questions concerning whether dialect familiarity predicted word recognition processing over and above the influence of relevant linguistic skills (vocabulary) and family background factors (maternal education).

Table 1
Demographic of participants and their scores across standardized assessments administered in Experiment 1

	AAE Listeners	GAE Listeners
n	21	35
Female	14	15
Mean Age (in months)	47.6 (7.8)	50.9 (5.7)
PPVT-4th Edition Mean SS (SD)	95 (10)	121 (17)
EVT-2nd Edition Mean SS (SD)	96 (12)	121 (15)
Maternal Education		
Low	17	2
Mid	2	6
High	1	26
Declined to report	1	1

Note. SS standards for standard score. SD stands for standard deviation.

Experimental lexical processing task

To investigate the effects of dialect familiarity on lexical processing, we used the visual world paradigm to measure children's real-time (online) recognition of real words. As is typical in the visual world paradigm, children saw four images in a 2 by 2 array while simultaneously hearing words in either AAE or GAE. In this task, children's gaze patterns to each of the images are automatically recorded allowing the evaluation of moment-to-moment processing as each word unfolds. Figure 1 provides an illustration of an example trial within the experimental paradigm.

Figure 1. An example trial from the spoken word recognition visual world paradigm.

Lexical items. All of the words in this study were monosyllabic nouns with an age of acquisition that ranged between 38.5 and 56.5 months and could be easily pictured by photographs from publicly available databases. All referents were paired with at least two photographs to maintain children's interest in the task. All noun-photograph pairs were normed in two preschool classrooms to ensure that children associated chosen photographs with intended lexical labels (see Appendix C for details about the visual stimuli and picture norming procedure).

The final stimuli included 24 unique target referents. Each target referent (*shirt*) was paired with three additional competitor items: a semantic (*dress*), phonological (*sheep*), unrelated (*bowl*) object. Appendix B provides a full list of the stimuli items included in this experiment.

Auditory stimuli. The auditory stimuli were recorded in a sound-treated room by two female speakers from Wisconsin and were edited using Praat (Boersma & Weenink, 2022). The GAE stimuli were recorded by a European American native speaker of GAE and the AAE stimuli were recorded by an African American speaker who had experience dialect-shifting between AAE and GAE. Speakers were instructed to use child-directed speech to produce all noun referents in two carrier phrases "*Find the* [noun]" or "*See the* [noun]". To capture linguistic differences between AAE and GAE, the production of carrier phrases and nouns retained phonetic and phonological properties that aligned with either AAE or GAE. For example, in the word *gift* the final consonant (i.e., [t]) was zero-marked in AAE (produced as [gɪf]). Further, to reduce anticipatory coarticulation effects between the transition from the carrier phrase to the target referent, all carrier phrases were recorded in the sentence "*Find the* egg" and target word recordings were cross-spliced with a period of 80ms of silence with the carrier phrase to create the final stimuli. Additionally, all of the stimuli were normalized to the same duration and intensity**.** Each speaker also recorded a set of reinforcing phrases in their native dialect. For example, speakers of AAE recorded the phrase *"You're doing so good"* in AAE which resulted in the following phrase *"You doing so good"*. These phrases were played at the end of each stimulus trial in their respective dialect conditions, meaning children heard "*You're doing so good*" for words in the block containing GAE productions and "*You doing so good*" in the block of eye-tracking containing AAE productions.

Procedure. The experimental task was designed in E-Prime ® Professional 2.0 (Psychology Software Tools, Inc 2010; Schneider, Eschman & Zuccolotto, 2002) and gaze patterns were monitored using the automatic Tobii T60XL eye-tracking system at a sampling rate of 60Hz. All children were introduced to the task under the pretense of playing an *I Spy* picture game. At the beginning of the experiment, children were instructed to look at the images and to listen carefully to each sentence. During the experimental trials, children saw four images on the Tobii monitor. Images remained on the screen for 2000ms to allow children to familiarize themselves with the pictures, then the critical auditory stimulus was presented, "*Find the gift.*"

At the end of the trial, an attention getting phrase (*"Look at that"*) was heard 1000ms after the target word offset and subsequent trials followed after the offset of the phrase. In addition to hearing motivational phrases, an animated puzzle was interspersed among the trials to increase children's participation in the task. Specifically, after children completed 6 or 7 trials, a few puzzle pieces that covered a child-friendly image were removed to reveal a portion of the image.

Children received two separate blocks of eye-tracking in which the auditory stimuli were grouped by dialect (i.e., children heard lexical items in AAE in one block and in GAE in a different block). All eye-tracking blocks were distributed across two separate visits to the lab, usually separated by a week. Further, the order of blocks was counterbalanced across participants and the trials within each block were randomized. This procedure was implemented to minimize potential order effects and to minimize opportunities for word recognition performance in one dialect to influence performance of listening to words in the other dialect.

Language Measures. In addition to the experimental task, children in this study were administered standardized measures of expressive and receptive vocabulary (*Expressive Vocabulary Test- 2nd Edition*, EVT-2, Williams, 2006; *Peabody Picture Vocabulary Test-4th Edition*, Dunn & Dunn, 2007, respectively). In the expressive vocabulary test, children were shown an image and asked to provide a label or a synonym for the pictured object. In the receptive vocabulary test, children were shown four images and were asked to point to the image that was named by an examiner. Each standardized assessment provides a standard score with a mean of 100 and a standard deviation of 15.

Data cleaning and preparation

Data cleaning and screening occurred in two phases. First, a customized script (provide Github link and citation) implemented a *deblinking* procedure which interpolated short windows of missing data, up to 150ms, if the child appeared to fixate on the same image before and after a *missing* window of data. The assumption of the interpolation process is that 150ms is too short a window for children to fixate to a different area of interest. The quality of the data was then examined in the time region of 250ms to 1750ms following the onset of the target word. All further data preparation including trial or participant exclusion and additional transformation of the data was implemented through the use of the eyetrackingR package (version 0.2.0; Forbes, Dink & Ferguson, 2021). Areas of interest (AOI), the quadrant in which one of the four images appeared, was defined by the boundaries of the four images on the display screen (450 by 450 pixel squares).

Gaze data exclusion. Trials were excluded if at least 50 percent of the data within a trial, during the analysis window, were missing. Additionally, participants were excluded if at least 50 percent of the trials in the block were not reliable for analysis (i.e., children needed to have at least 12 trials of reliable data in each block). Lastly, if children only completed one of the two Dialect conditions (e.g., Hearing AAE or Hearing GAE), the entire participant was excluded because performance across dialect conditions could not be reliably evaluated within subject. Listeners who were a part of the larger longitudinal study received two blocks of the lexical processing task in the dialect condition, Hearing GAE, and one block of eye-tracking for the condition, Hearing AAE. However, listeners who were not part of this project only received one block of eye-tracking in either dialect condition (Hearing AAE or GAE). To ensure comparable analytical conditions across all children, we analyzed only one block of eye-tracking in the GAE stimuli. For children who received two blocks in this dialect, we chose the block with the greatest amount of data.

Across participants, this resulted in 2,333 out of 2784 reliable trials across 56 participants. Two additional participants were tested, but these participants had greater than 50 percent missing data in at least one block of eye-tracking, and were thus excluded from all further analyses. Subsequent analysis of the data revealed that the proportion of missing data across participants did not differ across linguistic group membership (i.e., being an AAE or GAE Listener), $F(1, 2782) = .035$, $p < .85$ or the experimental condition, Dialect Heard (i.e., AAE or GAE), $F(1, 2782) = .084$, $p < .77$. The table below summarized descriptive statistics for the trials across linguistic groups before and after data cleaning.

Table 2
Summary of the track loss and sample analysis before trial-exclusion examined by listener group

Listener Group	**n**	**Number of Trials**	**Mean Samples**	**Percent Track Loss (%)**
AAE	21	1008	179	20.8
GAE	37	1776	179	20.9

Note. n is the number of participants across each listener group.

Table 3
Summary of the track loss and sample analysis after trial-exclusion examined by listener group

Listener Group	**n**	**Number of Trials**	**Mean Samples**	**Mean Percent Track Loss (%)**
AAE	21	858	179	12.9
GAE	35	1475	179	12.6

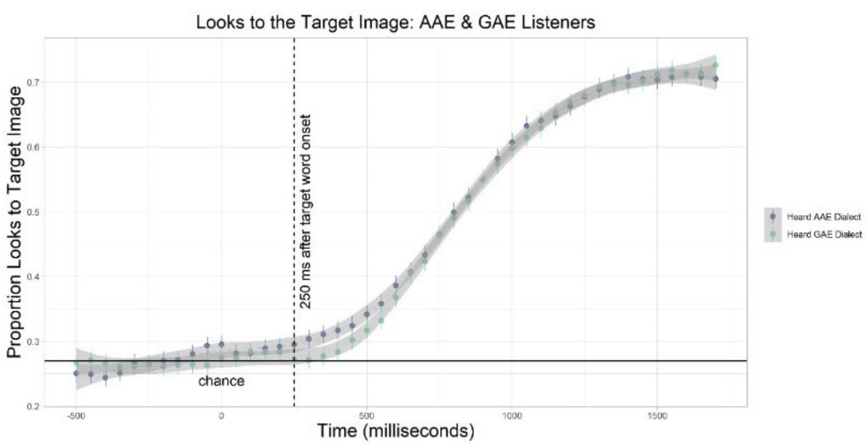
Note. n is the number of participants across each listener group.

Results

Following data cleaning and trial/participant removal, a generalized linear mixed effect model was used to examine the effect of dialect familiarity on word recognition while controlling for additional participant-level variables that have been previously shown to predict lexical processing accuracy, including maternal education and vocabulary size (Fernald et al., 2008).

The onset of the region of analysis was empirically determined by plotting the grand mean of the log-odds of looking at the target image for each Dialect Heard condition across both Listener Groups (i.e., AAE and GAE Listeners). The earliest time point at which the “fixation curves” departed consistently from chance occurred 250ms after the target word onset. The offset of the region of analysis was 1500ms after the starting point (~ 1750ms), a time window similar to what has been used in earlier studies on spoken word recognition (Chauncey et al. 2008; Fernald et al., 2008; Haith et al., 1993; Mahr et al. 2015; Marchman & Fernald, 2008; Swingley et al. 1999). Figure 2 depicts children’s proportion of looks at the target image across a larger window of analysis, 500ms prior to the onset of the target word to show that looks do not depart from chance until after 250ms after the target word onset

Figure 2. The grand mean in the proportion of looks to the target image beginning 500ms prior to the target word onset to 1750ms after the target word onset across both listener groups (AAE- and GAE-speaking children). The dashed, vertical line shows the point in time that is 250ms after the onset of the target word, which begins at 0ms on the graph below.



Generalized linear mixed effects analysis

The dependent variable was the log odds of looking to the target image and the predictor variables included fixed effects of Dialect Heard (AAE or GAE), Linguistic Group (AAE or

GAE speaker), Vocabulary (PPVT4 GSV score as a centered continuous variable), and Maternal Education Group (Low, Mid, High). Additionally, two-way interactions between Dialect Heard and Linguistic Group as well as Vocabulary and Maternal Education were fit. Lastly, the model included a by-participant random intercept and random slope for Dialect Heard. A more parsimonious model excluding trial-level random effects was fit due to significant issues with model convergence. The R syntax is provided in Table 4 with the summary of the model fit.

A mixed effects logistic regression in the region of analysis (i.e., 250ms to 1750ms relative to the target onset) was fit to the data. Across this region, there was a significant effect of vocabulary and a significant interaction between vocabulary and maternal education for children from families with mid and high levels of maternal education. The results of this model suggest that the log odds of looking to the target image is best predicted by vocabulary size. Additionally, the significant interaction indicated that the effect of vocabulary was more pronounced for children from families with lower levels of maternal education as compared to children with higher levels of maternal education (i.e., some college or more), as illustrated in Figure 3. Fixed effects of Dialect Heard, Linguistic Group, Maternal Education, and all other interactions were not significant. Moreover, the best fitting model for the data did not even include Dialect Heard and indicated that Vocabulary was again the strongest and best predictor of lexical processing.

Table 4.
Summary of parameter estimates for the mixed-effects logistic regression model fit to the gaze data. These models were fit using maximium likelihood.

R Syntax: cbind (LooksTarg, LooksNotTarg) ~ Dialect Heard*Linguistic Group + Vocabulary*Maternal Education Groups + (Dialect Heard | Participant)

Fixed Effects

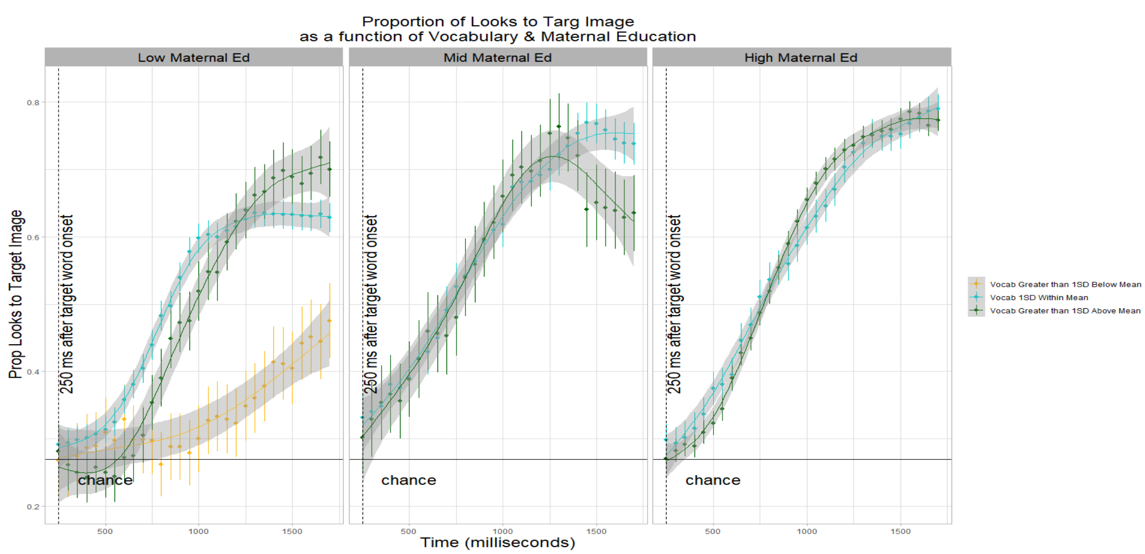
	b	**SE**	**z**
Constant	.31	.11	2.76 **
Dialect Heard	-.08	.08	.31
Linguistic Group	-.18	.15	.22
Vocabulary (PPVT-4 GSV)	.39	.09	4.21 ***
Maternal Ed (Mid)	.13	.16	.83
Maternal Ed (High)	.12	.16	.76
Dialect Heard*Linguistic Group	.08	.10	.76
Vocabulary * Maternal Education (Mid)	-.24	.12	-2.05 *
Vocabulary * Maternal Education (High)	-.26	.12	-2.04 *

Random Effects

Parameter	**Variance**	**SD**	**Correlations**
Constant	.11	.33	
Dialect Heard	.12	.34	-.52

p <.001***, *p* < .01 **, *p* < .05*

Figure 3. Mean looks to the target image for three vocabulary groups as a function of maternal education level



Discussion

To summarize, the results of Experiment 1 suggest that GAE and AAE Listeners, as young as age 3, demonstrate remarkable flexibility in recognizing familiar words that contain linguistic differences between AAE and GAE. Additionally, this study found that vocabulary was the most reliable predictor of word recognition across linguistic communities. This is the first study to use the visual world paradigm to investigate the relationship between dialect familiarity and spoken word recognition among preschool speakers of either AAE or GAE. Earlier studies examining the influence of speakers' use of AAE or GAE and lexical processing have focused on children's sensitivity to typical pronunciation and mispronunciation of real words in children's native dialect (Law et al. 2017), but did not assess lexical processing in the less familiar dialect.

Overall, the findings from Experiment 1 suggest that dialect variation does not influence spoken word recognition in this very simple task. First, we found neither a main effect of dialect (as measured by differences in spoken word recognition when hearing either GAE or AAE) nor an interaction between dialect and linguistic group. Second, the best fitting model of the data excluded dialect familiarity from the level 1 model structure entirely. This result is not necessarily in conflict with a perceptual analysis hypothesis, given the demands of the task itself and the dialect differences contained in the stimuli. First, the demands of the task were relatively simple (i.e., listeners had to recognize highly familiar words in simple sentences such as *find the X*). Second, differences between the two dialects were limited to phonology and were not observed for all target words. For example, the lexical referent *gift* can be produced as either [gɪf] or [gɪft] in AAE depending on the surrounding phonological environment, but words like *fly* are not pronounced differently in AAE or GAE. Linguistic differences between AAE and GAE may be more attenuated at the level of words, which facilitated children's flexible adaptation in this study.

Evidence from this study converges with the results reported by Edwards et al. 2004. These authors found that AAE-speaking children's comprehension of GAE was more accurate for words that did not contain phonological or morphological ambiguities as opposed to words that contained ambiguities. It could be the case that the words chosen for this study did not provide many opportunities words for lexical uncertainty, which resulted in equally efficient processing of words in the familiar and less familiar dialect of English. Moreover, while it is the case that carrier phrases preceding the target word contained contrastive features between the dialect (e.g., zero-marking of the final consonant in the cluster of *find* and naturalistic prosodic features of AAE) this may not have provided sufficient context to highlight linguistic differences.

Relatedly, prior research has also suggested that the magnitude of difference between linguistic varieties may be an important factor for observing effects of linguistic mismatch (Nathan et al., 1998; Barker & Turner, 2015). It is plausible that effects of dialect familiarity among dialects like AAE and GAE, will be observed only when linguistic differences are more pronounced. At the level of words, there are few opportunities to encounter salient differences between AAE and GAE. It could be the case dialect variation effects will be observed when the auditory stimuli capture a broader and more naturalistic sample of the range of linguistic differences between AAE and GAE. This is crucial because longer utterances allow for greater differences across dialects, and therefore create more opportunities to encounter lexical uncertainty or ambiguity among lexical items. African American English is typically described as a linguistic system containing prosodic, phonological, and morphosyntactic patterns and rules that are both shared and distinctive with other linguistic varieties of English (Green, 2002, 2010; Wolfram & Schilling, 2015). Additionally, some researchers have even suggested that it is a variety that is primarily characterized by differences in morphosyntax (Washington et al. 2018; Washington & Craig 1998). Experiment 1 examined prosodic and phonological patterns, but did not include morphosyntactic patterns which make up a large part of the inventory of differences that exist between AAE and GAE. It could be the case that experimental designs in which linguistic differences are more apparent are necessary to observe effects of dialect familiarity on spoken language comprehension.

As noted above, simple spoken word recognition may be too simple of a task to observe effects of dialect variation. In this experiment, children had to recognize highly familiar words produced in very simple phrases. Therefore, even if dialect differences were sufficiently salient, perhaps recognizing words that are well known by most of the children did not provide a demanding enough linguistic environment for effects of dialect variation to be pronounced. This would indicate that future research on dialect variation should employ experimental paradigms in which there are greater linguistic demands during real-time spoken language processing.

Consistent with previous studies evaluating relationships between vocabulary and lexical processing (Fernald, Marchman, & Weisleder, 2013; Fernald, Swingley, & Pinto, 2001; Hurtado et al., 2014; Law & Edwards, 2015)**,** this study found that vocabulary was a strong predictor of lexical processing in children. Children with larger vocabularies are likely to more quickly and accurately fixate to the target referent in this task compared to children with smaller vocabularies. This finding converges with previous results reported by Law and Edwards (2015) who found that vocabulary size was a significant predictor of lexical processing in children ages 30 to 57 months; children with larger vocabularies exhibited longer and more rapid fixations to the target image. Interestingly, we observe similar findings in spite of using different measures of vocabulary. Law et al. used a measure of expressive vocabulary (i.e., *Expressive Vocabulary Test, 2nd Edition*) whereas this study utilized a measure of receptive [1]vocabulary (i.e., *Peabody Picture Vocabulary Test, 4th Edition*). Furthermore, the results of this study provide tentative support that vocabulary is a particularly sensitive predictor of children's word recognition if they are from families with lower levels of maternal education, a measure of socioeconomic status. This result suggests that vocabulary growth is very important among children with smaller vocabularies which then impacts their lexical processing to a greater degree. This might differ for children with relatively large vocabularies. It is plausible that once these children have acquired a certain number of words, vocabulary no longer serves as a sensitive predictor of lexical processing efficiency of highly familiar words.

In summary, this study sought to understand two important research questions. First, it evaluated the relationship between dialect variation and real-time spoken word recognition through a perceptual analysis lens. Second, this study explored the relationship among maternal education, children's expanding linguistic skills, and dialect variation on lexical processing. The central finding of Experiment 1 suggests that 3- to 5-year-old AAE- or GAE-speaking children recognized words in either AAE or GAE. Similar to the results of other studies (Fernald et al. 2013; Hurtado et al. 2014; Law et al. 2015), we found that vocabulary facilitates lexical processing. Moreover, this result was most relevant for children from families with lower levels of maternal education.

[1] We also evaluated a model replacing the measure of receptive vocabulary with a measure of expressive vocabulary, and the results we obtained were similar. The critical difference is that the interaction between maternal education and vocabulary is absent for children from families with mid levels of education (i.e., some college or trade school) but remains significant for the interaction for children with high levels of maternal education (i.e., four-year college degree of higher).

Further, we hypothesized that finding no effect of dialect could be attributed to either the demands of the task (i.e., recognizing highly familiar words) and/or the magnitude of difference observed across dialects (i.e., minimal differences due to either prosodic or phonological differences). Therefore, the question that remains is whether effects of dialect variation are observed when the demands of the task are increased and differences across the dialects are more salient or pronounced. This question is crucial if we are to understand the effects of dialect variation on spoken language processing and if we aim to examine the limits of the perceptual analysis hypothesis in adequately explaining relationships between dialect variation and spoken language comprehension.

Experiment 2

To determine the effects of task complexity and dialect salience on spoken language processing of linguistic differences between AAE and GAE, we conducted a second experiment that examines these factors more directly. Specifically, Experiment 2 evaluates the influence of dialect familiarity on children's use of semantic prediction during sentence processing. We address the limitations of Experiment 1 in two ways. First, the stimuli capture a broader range of phonological and morphosyntactic differences across dialects (i.e., enhancing dialect salience). Second, we examine children's ability to leverage semantically facilitating information across sentences, a task that is more complex than recognizing familiar words in very simple phrases (i.e., increasing task complexity).

One reason listeners process spoken language so quickly and accurately is that they are able to use both lower-level (e.g., acoustic-phonetic and phonological information) and higher-level cues (e.g., conversation topics or constraining lexical-semantic information) to predict upcoming sounds and words. One higher-level cue that is available early in language acquisition is the listener's use of lexical-semantic information. Several studies have shown that children as young as age 2 reliably use semantic information to predict upcoming words. Using the looking-while-listening paradigm, Fernald et al. (2008) found that 26-month-old children generated anticipatory looks to a target object (e.g., *cake*) when hearing a semantically facilitating verb (e.g., *eat*) even before the target object occurred in the speech signal and similar findings have been reported by Mani & Huettig (2012) and Nation et al. (2003). Additionally, other studies have shown that listeners anticipate thematically appropriate arguments when given an *agent + action* and even predict final noun referents when provided with predictive verb contexts (Kamide et al. 2003; Mani & Huettig, 2012). Overall, these studies suggest that the use of at least some lexical-semantic cues for prediction,

Much of our understanding about how listeners integrate lexical-semantic cues when listening to unfamiliar dialects comes from research with adults. Clopper (2012) examined semantic prediction among adult participants and found that they showed decreased semantic prediction when processing unfamiliar regional dialects of English. In a familiar or unfamiliar regional dialect, participants listened to sentences with a missing final word that contrasted in semantic predictability, "The judge is sitting on the _____" (*high*) vs. "I'm talking about the ____" (*low*). Listeners were instructed to generate final words (e.g., bench) that best fit the sentential context. This study found that lexical-semantic cues in the highly predictive contexts were most beneficial to listeners in the familiar dialect condition but not the unfamiliar dialect condition.

However, we know far less about how children contend with linguistic variation while integrating multiple sources of information. Children are at period of development in which their linguistic knowledge is developing and maturing as a result of different intrinsic, environmental, and experiential factors. We hypothesize that even early acquired linguistic processes like semantic prediction will be affected by language variation, and may influence children to a greater degree than adults.

To date, there has only been one study to evaluate relationships between linguistic variation and semantic prediction in children. Holt and Bent (2017) examined 5- to 7-year-old children's ability to use semantic context to support word recognition of foreign-accented speech in an adverse listening condition (i.e., multitalker babble). Children listened to sentences in either Mandarin-accented English (non-native) or American English (native) and repeated sentences which contained final words that were either highly predictable or less predictable from the sentence context. The authors observed main effects of context and accent familiarity, such that word recognition scores were higher in high predictability contexts and children were more accurate in the native-accent category. However, they did not observe an interaction between semantic context and accent familiarity, meaning that semantic context was equally beneficial in both accent conditions. A limitation of this study is that the experimental conditions that were examined may have created an extremely adverse listening environment. In their study, children had to contend with two sources of signal degradation, noise and accent variation. While it is important to capture naturalistic experimental contexts in which listeners are hearing dialects in less than optimal environments as in the case of noisy environments like classrooms, it is plausible that noise masked the ability to examine the contribution of accent variation. This might be especially true since the children performed poorly across all accent conditions (familiar and unfamiliar), as evidenced by low scores across the age range. Hence, including two different sources of signal degradation, at least for studies with children, makes it harder to disentangle effects of accent variation on semantic prediction.

Additionally, this study employed a blocked design in which children listened to a set of sentences varying in context predictability in the non-native accent separately from the native accent. Previous work by Clopper (2012) showed that the interactive effect of context and dialect familiarity was maximized when individuals were presented with dialects in an unblocked arrangement, meaning that the unfamiliar dialect stimuli were interspersed with the familiar dialect stimuli. To reliably disentangle effects of dialect variation, it may prove useful to design experiments where effects of variation are more salient or at the very least less confounded by additional sources of variation (or signal degradation). Furthermore, if the goal is to understand the validity of the perceptual analysis hypothesis in explaining effects of dialect variation on spoken language comprehension, then studies that quantify effects of dialect separately from other sources of signal degradation may be useful.

Hence, the primary research question in Experiment 2 addresses whether dialect variation influences semantic prediction but particularly among school-age children. As described earlier, we evaluate this question by using a more demanding linguistic task (i.e., semantic prediction) and increasing the salience of dialect differences between AAE and GAE. Additionally, the validity of the perceptual analysis framework is once more tested for its ability to adequately explain relationships between children's spoken language processing and dialect familiarity.

In this study, the predictions derived from a perceptual analysis framework in Experiment 2 differ slightly from Experiment 1 because there are two important factors to consider: a) the salience of linguistic differences between the dialects and b) the availability of facilitating linguistic cues. The first factor is the most intuitive, that being that effects of dialect variation will be enhanced in this study because the linguistic differences between AAE and GAE are more prominent (longer sentences with multiple morphosyntactic and phonological differences between the dialects). The second factor to consider is the sentential context. Previous studies on accent familiarity have suggested that supportive linguistic contexts (i.e., higher-level cues), and more specifically, semantically predictive linguistic information, facilitate children's processing of less familiar accents (Clopper, 2008; Creel, 2012). The results of these studies suggest that semantically predictive information can be leveraged by listeners creating fewer opportunities for less familiar accents (and potentially dialects) to disrupt spoken language processing. This would suggest that the effects of dialect differences may be minimized when supportive or predictive context cues are available, but maximized when these cues are less available.

Given these two sources of information (i.e., dialect and context), the following predictions are anticipated from a perceptual analysis account:

Prediction 1: A main effect of dialect will be observed: children will have quicker and more reliable fixations to the target in a familiar dialect relative to a less familiar dialect.

Prediction 2: An interaction between dialect familiarity and semantic prediction will be observed. Based on previous research, we predict that effects of dialect variation will be greater in contexts where facilitating semantic information is not available (i.e., neutral verb contexts).

Methods

Participants

Listeners

Thirty-five children from areas near Washington, D.C., Maryland, and Virginia (DMV) in the United States participated in this study. Four additional children were recruited but were not included in the final sample for the following reasons: a) their caregivers/legal guardians reported that children were receiving special education services or had a developmental delay (n= 1), b) children did not complete the eye-tracking experiment due to lack of compliance or equipment failure (n = 2), or c) children scored more than 1.5SD below the mean on standardized measures of language or on a measure of non-verbal intelligence (n = 1). Table 5 provides additional details about the participants in the study.

Data collection began two months prior to the start of COVID-19. At the onset of the project, families were contacted through a combination of online and in-person recruitment efforts. The majority of children were recruited from a University-based consortium of families who expressed interest in participating in child development research. Families from this database were provided information about the study and invited to participate through email. Children were also recruited from a private school in Southern Maryland where the majority of children were African American (i.e., the school had an attendance rate of 99% African American students). Families from the private school were invited to attend an afterschool presentation about literacy and language development. At the end of the presentation, the first author provided information about the study and ways to participate. This second approach to recruitment was used to include children with a broad range of dialect use in this region. Despite extensive efforts to collect a larger sample of children who spoke AAE, the pandemic severely disrupted in-person data collection. Two years after the start of COVID-19, in-person testing resumed but few parents reported feeling comfortable with in-person testing. By implementing several hygiene-related protocols to address the concerns of caregivers and ensure the safety of families, we were able to recruit three additional participants who demonstrated a broader range of dialect variation patterns. However, the sample sizes remained relatively small (n=12).

Table 5
Demographic of participants and their scores across standardized assessments administered in Experiment 2

	GAE Listeners	**NMAE Listeners**
n	23	12
Female	14	6
Mean Age (in months)	93.4 (13.8) *range=72-119*	102.9 (15.3) *range=76-119*
PPVT-4th Edition *Mean Standard Score* (SD)	**119 (17)** *range=92-153*	**107 (18)** *range=83-147*
DELV-ST *Dialect Density Score*	.11 (.12) *range=.00-.38*	.40 (.26) *range=0.0-1.00*
Maternal Education		
Low	2	6
High	21	3
Declined to report	NA	3
Racial Background		
African American	9	10
European American	14	2

Dialect group assignment among listeners

The native dialect of all listeners was confirmed using a norm-referenced assessment of dialect use, *The Diagnostic Evaluation of Language Variation, Part I, Screener* (DELV-ST, Seymour, Roeper, & de Villiers, 2003). This norm-referenced test of dialect variation examines children's responses to 15 items that were designed to elicit common phonological and morphosyntactic patterns aligning with nonmainstream dialects of English that provides an age-normalized criterion score with three levels: a) a speaker of Mainstream American English, b) a speaker producing *some* variation from Mainstream American English or c) a speaker producing *strong* variation from Mainstream American English. Children who received a categorical score of Mainstream American English were assigned to the GAE listener group. Children who received a score of Some or Strong variation were assigned to the nonmainstream American English listener group. We use the term nonmainstream American English because we had a diverse group of children who scored in this category including confirmed speakers of AAE and children who produced nonmainstream American English (NMAE) patterns but it was not certain whether these forms were AAE or another nonmainstream variety of English in the DMV locale. In the final sample, 23 children were identified as GAE listeners (mean=102.3 months; 14 female; 9 African American and 14 European American) and 12 children were identified as NMAE listeners (mean=100.7 months; 6 female;10 African American and 2 European American).

Speakers

The sentences were recorded by two 21-to 25-year-old female speakers who lived near College Park, Maryland at the time of the study. The speaker of General American English was born and raised in Maryland and self-reported being a speaker of only GAE throughout her childhood and adulthood. The speaker of African American English self-reported being a proficient bidialectal speaker of AAE and GAE. This was also confirmed by the first author during an informational interview in which dialect-shifting was observed throughout the duration of the conversation. Some of the features produced during the course of the conversation included use of habitual *be*, optional use of copula, and past tense subject verb agreement (was/were). The bidialectal speaker reported that they were raised in Detroit, Michigan but spent many adolescent and adult years with family nearby Maryland.[2]

Auditory Stimuli

[2] The author recognizes that AAE is not a single dialect but a family of dialects and differs across generations and regions, which further influences a speaker's production of AAE. This study focused on features that have been reported as the most widespread and stable among speakers of AAE (Craig & Washington, 2004; Green, 2010).

The experimental sentence processing task consisted of a total of 36 sentences. Of these sentences, 24 sentences were critical trials. Among the critical trials, half of the sentences included a predictive verb (e.g., *Every day, Caleb helps his friend* ***read*** *the book*) and the other half included a neutral verb (e.g., *Every day, Caleb helps his friend* ***choose*** *the book)*. Each sentence containing a predictive verb was matched to a neutral verb. In other words, among the paired trials *(read book vs. choose book)*, the preamble of each sentence was the same, but the verb differed so that sentences contained either a predictive or neutral verb. In addition to semantic predictability differences, we examined the effect of dialect by creating two versions of each sentence type. Phonological and morphosyntactic differences between AAE and GAE were embedded into each sentence. In the example sentence mentioned above, the following features were embedded: a) *invariant be,* b) *alveolarization,* and c) *zero-marking of final consonants*. The inclusion of these AAE features resulted in the following sentence, (*Caleb be helpin' his frien'* *read/choose the book;* see Table B2 for the complete stimuli list).

Additionally, for each sentence, the predictive verbs and their corresponding nouns (i.e., target referents) were normed prior to the experiment using a cloze sentence task that was presented to participants on Amazon Mechanical Turk (*see Supplementary Materials for details).* This procedure ensured that all target referents (**book**), in the critical trials, were adequately predicted by preceding predictive verbs (**read**) and perceived as admissible when preceded by neutral verbs (**choose**). An additional 12 filler trials were included. Each filler sentence included a neutral verb (**see**) that corresponded to a referent (**jewel**) that differed from the referents included in the critical verb-noun pairs. Filler trials did not include predictive verbs and were included to promote visual attention to the other images on the screen.

All auditory stimuli were recorded using a Shure SM51 microphone in a sound-attenuated booth. 1000ms of silence was added to the beginning of the auditory stimuli. The sentences were normalized for duration across dialect conditions through Praat (Boersma & Weenink, 2022) by either lengthening or shortening the duration of the segment normalized. The mean sentence length including the period of silence at the start of the audio file was 3546ms (sd = 111ms). The critical verb onset was time-locked to begin at 2815ms and the noun was time-locked to begin at 3200ms.

Lexical items and visual stimuli. All of the words in this study were monosyllabic and bisyllabic nouns with an age of acquisition that ranged between 33.6 and 74.6 months (mean = 51.8 months, sd = 11.2 months) and could be easily pictured by images from publicly available databases (Morrison, Chappell & Ellis, 1997; Dunn & Dunn, 2007). We selected four images for each referent because children would see each object across multiple trials (e.g., an image of *book* was presented four times, once during the predictive trial, neutral trial, and two filler trials in which a foil referent was heard but an image of book was present in the display). Images of the lexical items were normed in two kindergarten classrooms at a school, the Center for Young Children which is affiliated with the University. (*See Appendix C for visual stimuli details and norming procedures)*.

Procedure. The experimental task was designed using the SR Research Experiment Builder® software. Gaze patterns were monitored using either the automatic SR Research arm-mounted Eyelink 1000 Plus (n=17) or the Portable Duo eye-tracking system (n=18). Two separate eye-tracking systems were used because children were either tested at the University of Maryland in an office in the laboratory or at an offsite school location in a quiet room. The Portable Duo provides the same level of precision, accuracy, and data loss concerns as the Arm-

Mounted eye-tracker, but it is portable and provided an opportunity for us to evaluate a broader sample of children including speakers of nonmainstream speakers of English. This level of portability was necessary because we recruited children from a variety of locations, some of which were easily accessible to families and others which were less accessible due to distance and public transportation access.

The same procedure as described in Experiment 1 was used to introduce children to the visual world paradigm. However, in contrast to Experiment 1, at beginning of each trial, children were shown the four images in the display for 1200ms so that they could familiarize themselves with the images on the screen. Following this period, a fixation attention-getting video (e.g., a bullseye) appeared in the center of the screen. Children's fixation to the attention getter for at least 500ms initiated the onset of the critical auditory stimulus. Children were instructed to listen to the entire sentence and then select the picture that matched what they heard using an electronic stylus to tap one of the four images on the touchscreen monitor. The trial terminated after children selected an image and a new trial was displayed. This designed permitted researchers to assess children's accuracy in word selection and gaze patterns during the duration of the auditory stimuli.

A within-subjects design was used. Children heard sentences that contrasted in semantic predictability in two dialects of English, AAE and GAE. All of the trials were pseudorandomized offline and then administered to children (see Supplementary Materials for further information about the pseudo-randomization). In total, listeners heard 72 sentences, 48 of which were critical trials. The stimuli were distributed across two separate lists; a single list constituted one block of eye-tracking. Children were administered the two blocks of eye-tracking in the same testing session, but separated by a short break. The break consisted of a short game with the examiner and the completion of one of the standardized/norm-referenced assessments. Similar to Experiment 1, to minimize effects of task order, the order of the lists presented were counterbalanced across all listeners (e.g., some listeners received List 2 in their first block, while other received List 1 in their first block).

Language Measures. Children were administered three standardized tests to evaluate their language skills and nonverbal intelligence. Receptive vocabulary was assessed using the *Peabody Picture Vocabulary Test- 4th Edition* (PPVT-4; Dunn & Dunn, 2007). This standardized test provides a standard score of 100 with a standard deviation of 15. Children's nonmainstream dialect was assessed using Part 1 of the *Diagnostic Evaluation of Language Variation-Screening Test* (DELV-ST; Seymour, Roeper, de Villiers, 2003). The first part of this screener provides a categorical index of whether children speak a dialect that varies strongly from GAE, varies somewhat from GAE, or aligns with GAE. In this study, we extracted two measures of children's dialect use, a criterion score (i.e., the categorical index score provided by the DELV-ST) and a quantitative measure of dialect density (i.e.¸ feature rate usage that is calculated by dividing the number of features aligning with NMAE by the total number of GAE and NMAE aligning responses). This second measure is commonly used in research on dialect variation to assess how use or production of AAE impacts measures of academic achievement (Gatlin & Wanzek, 2015; Puranik et al., 2020, NP Terry et al., 2016).

Eye-tracking data preparation for gaze data analyses

Data cleaning and screening followed the same procedure described in Experiment 1. Because trials across experimental conditions were intermixed within each block of eye-tracking, children were excluded from all further analyses if they did not have an adequate amount of data in at least one block of eye-tracking (50% of reliable data available in at least one block of eye-tracking). This procedure was implemented to retain as much data as possible. The quality of the data was then examined in the time region of 2800ms to 5500ms, the region beginning at the onset of the critical verb (read/choose) to 2000ms after the onset of the target referent (book).

Across the participants this resulted in 1337 out of 1638 reliable trials. Two additional children were tested, but they had greater than 50% of missing data in more than one block of eye-tracking and were thus excluded from all subsequent analyses. Further analysis of the data revealed that the proportion of missing data across participants did not differ across the experimental condition, Dialect Heard (i.e., AAE or GAE), $F(1, 1335) = .046$, $p = .83$, or the experimental condition, semantic predictability (i.e., predictive or neutral verbs), $F(1, 1335) = .19$, $p = .65$. The tables below provide a brief summary of the track loss analysis for participants belonging to each linguistic group before and after data cleaning. Additionally, all eye-tracking analyses were restricted to correct responses.

Table 6

Summary of the track loss and sample analysis before trial-exclusion examined by listener group

Listener Group	**Number of Participants**	**Number of Trials**	**Mean Samples**	**Mean Percent Track Loss (%)**
GAE Listeners	23	1108	2952	35.9
NMAE Listeners	12	530	4178	35.7

Table 7

Summary of the track loss and sample analysis after trial-exclusion examined by listener group

Listener Group	**Number of Participants**	**Number of Trials**	**Mean Samples**	**Mean Percent Track Loss (%)**
GAE Listeners	23	885	3117	27.1
NMAE Listeners	12	452	4217	30.5

Analyses & Results

Three separate analyses are reported below. The first analysis evaluates the effect of dialect familiarity and semantic prediction on children's accuracy in selecting the target referent. The second analysis uses a cluster-based permutation method to identify a region of analysis for the effect of semantic prediction. Lastly, the third analysis investigated the influence of dialect familiarity, semantic prediction, and other language related predictors on spoken language comprehension in the window of analysis identified from second analysis.

Response accuracy. Overall accuracy in this task was above 85% across all experimental conditions. A generalized linear mixed effects model with response accuracy coded as 0 or 1 was used to examine whether accuracy was influenced by Dialect Heard and Semantic Predictability. The model included fixed effects for Vocabulary, Age, Dialect Heard, and Semantic Predictability. A two-way interaction between Dialect Heard and Semantic Predictability was also included. Lastly, a random intercept for participant was included. All continuous variables were mean-centered prior to the analysis. Neither the main effects nor interactions of any of the included predictors significantly predicted response accuracy (p's $> .05$). Table 8 contains the model estimates from this analysis and Figure 4 illustrates mean accuracy across conditions.

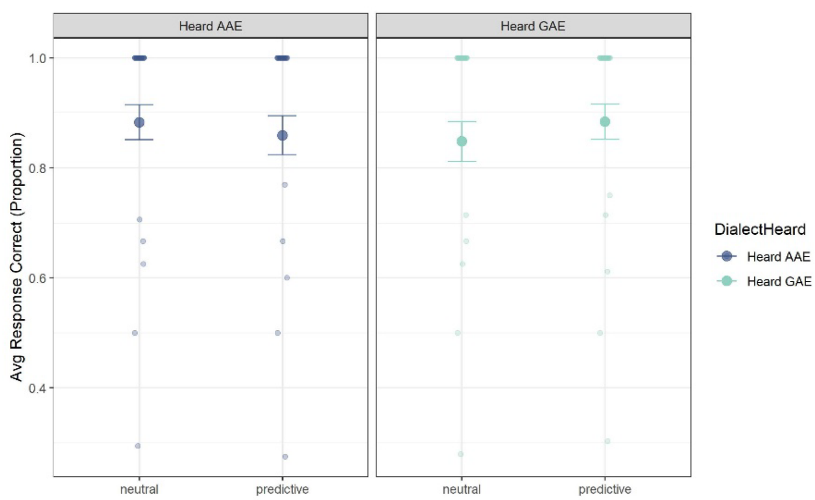
Table 8
Model summary for the response accuracy analysis.

*R syntax: Response Accuracy ~ Dialect Heard*Semantic Condition + Vocabulary + Age + (1|Participant)*

Fixed Effects	***b***	**SE**	**z**
Constant	11.48	2.09	5.48 ***
Dialect Heard (GAE)	-.04	.30	-.14
Semantic Condition (predictive)	-.05	.31	-.16
Vocabulary (GSV)	-.29	1.79	-.16
Age	.63	1.88	.33
Semantic Condition x Dialect Heard	.12	.43	.27
Random Effects	**Parameter**	**Variance**	**SD**
	Intercept	101.7	10.09

*p < .001 ****

Figure 4. Response accuracy as a function of Dialect Heard Condition (AAE or GAE) and each Semantic Condition (Predictive or Neutral Verbs).



Cluster-based permutation analysis. Cluster-based permutation analysis, originally developed for analyzing EEG data, is a non-parametric method that is used to delineate contiguous clusters of time in which statistical effects can be reliably observed. Permutation testing is a useful approach to analyzing eye-tracking data when researchers are uncertain about the windows of time in which experimental condition effects should be observed. For example, traditional approaches to analyzing eye-tracking data typically entail the examination of interactions between experimental condition effects and predetermined time regions or time derived measures (e.g., time-derived polynomials as is used in growth curve analysis). These approaches seem relatively robust when researchers evaluate looking patterns among participants who either belong to well-studied linguistic populations or provide sufficient amounts of data to estimate relevant time windows for condition effects. However, these approaches may be less reliable when studying linguistic communities who are less represented in the literature or when there are theoretical uncertainties about when effects should occur especially if there are departures from the methodologies implemented in prior studies. These limitations, among others, are often present in online processing studies with children or individuals form linguistic communities who are less well represented in the literature. Additionally, our experiment included sentences with longer preambles than previous studies (Borovsky et al., 2012; Mani & Huettig, 2012) to make dialect differences more salient. For these reasons, we use an empirical approach to determine the windows of interest for evaluating effects of variation on spoken language comprehension. Cluster-based permutation tests provide help to resolve some of these limitations by analyzing separate bins of data for significance, creating clusters among statistically significance adjacent windows, and testing it against an empirically derived null distribution while correcting for multiple comparisons. A general description of cluster-based permutation can be found in Maris and Oostenveld (2007). However, in this paper I provide a detailed description of the procedure used by the authors to carry out this analysis.

Procedures. Prior to implementing the cluster analysis, the data were filtered to the following region: 2800ms to 5500ms (the region from the onset of the verb to approximately

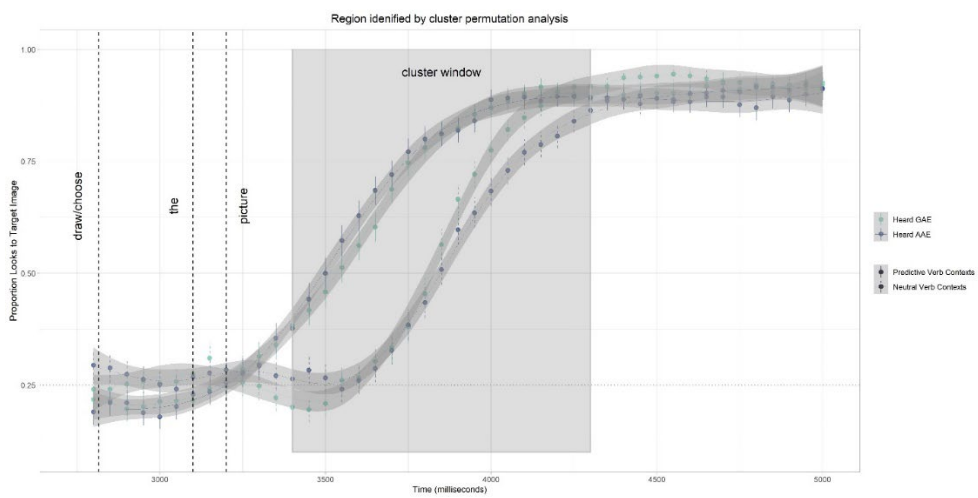
2000ms after the target word onset). This general region aligns with both prior evidence (Mani & Huettig, 2012; Blomquist et al., 2021, Borovsky et al., 2012) and theoretical predictions which suggest that listeners should exhibit anticipatory looks to the target image even prior to the onset of the noun. We then used eyetrackingR to create 50ms time bins, which provided some smoothing to the data removing extraneous sources of noise which may have obscured conditional effects or trends among other variables in the data set.

Following this initial filtering procedure, the *permutes* package was used to implement the cluster analysis which examined regions of time in which main effects of semantic prediction would be observed across linguistic groups. The permutes package was specifically designed to carry out permutation testing as described by (Maris & Oostenvald, 2007). Its procedure identifies time points where significance can be found and supports mixed-effects models using an implementation approach of Lee and Braun (2012). We used this package (Version 2.6; Voeten, 2022) to fit a mixed effects model predicting the log odds of looking to the target image as a function of a fixed effect for Semantic Condition, a by-subject random intercept and slope for Semantic Condition, and a by-item random intercept.

This package runs a permutation test at each time bin, correcting for multiple comparisons, identifying locations in which the experimental condition effect began and ended. Adjacent time bins with a significant test statistic at an alpha level of .05 are clustered together, indicating specific clusters of time in which looks to the target image were likely to be influenced by the critical predictor (i.e., semantic prediction). Due to issues with convergence and the large number of analyses across permutations we did not include random intercepts for item or any other fixed effects (e.g., Vocabulary, Age, or Dialect Heard). These additional variables were evaluated in a separate logistic mixed effects model which is reported below.

The cluster analysis identified a cluster of eighteen contiguous 50ms bins (i.e., 585-1485ms after the onset of the predictive verb) in which the effect of Semantic Condition met the criterion threshold for significance (a *p*-value that is smaller than the criterion level, a = .05). Figure 5 depicts the window of time that was identified by the cluster analysis.

Figure 5. The window of analysis selected by cluster permutation testing across all children.



Generalized linear mixed effects model. In the final analysis, a generalized linear mixed effects model was fit to the cluster window (see Figure 5) to examine the influence of dialect variation, semantic condition, and additional language measures on fixation to the target image among all listeners collapsed across the listener groups. Because of the unequal sized listener groups that were included in the sample (23 GAE Listeners and 12 AAE Listeners), listener group was not included as a predictor variable. Instead, the continuous measure of dialect density was used as a predictor. Like Experiment 1, the dependent variable was the log odds of looking at the target image. The predictor variables included fixed effects for Semantic Condition (Predictive or Neutral), Dialect Density, Dialect Heard (GAE or AAE), Vocabulary

(PPVT4 growth scale values), Age (in months), and Race. It also included two-way interactions between Dialect Heard and Semantic Condition, Vocabulary and Semantic Condition, Dialect Heard and Vocabulary, Dialect Density and Semantic Condition, Dialect Density and Dialect Heard. The model also included two 3-way interactions, an interaction among Semantic Condition, Dialect Heard, and Vocabulary and an interaction among Semantic Condition, Dialect Density, and Dialect Heard. Lastly, the model included a participant-level random intercept and random slope for the effects of Dialect Heard and Semantic Condition and an item-level random intercept and random slope for Semantic Condition. Prior to fitting the model, all continuous variables were centered and categorical variables were contrasted coded using the binary codes -.5 and .5.

Contrast coding was used to code the categorical variables (Davis, 2010) rather than dummy coding to aid the interpretation of effects within the model, because it estimates several interactions involving multiple categorical variables. In this model, the intercept represents the grand mean of all of the variables within this data set as opposed to the mean of a particular level within a category. In other words, the intercept is the average log odds of looking at the target image across Semantic Conditions and Dialects Heard for individuals of average age, dialect density, and vocabulary. The contrast corresponding to Semantic Condition evaluates a change in log odds of looking at the target image for neutral sentence contexts from a reference level (e.g., predictive sentences). The contrast corresponding to Dialect Heard reflects a change in log odds of looking at the target image when hearing AAE from a reference level of hearing GAE. Lastly, the contrast corresponding to Race examines the change in log odds of looking at the target image for a European American child as compared to an African American child. Race was included as a predictor because it is possible that the African American listeners had more experience with AAE than European American listeners, even if this wasn't exhibited in their productions on the DELV-ST. The R syntax and summary of the model estimates are included in Table 9.

There was a significant main effect of Semantic Prediction, meaning that children were more likely to fixate to the target image (*book*) when hearing a predictive verb (*read*) as opposed to a neutral verb (choose). There was also a significant main effect of Age, indicating that overall looks to the target image improved as age increased. A significant interaction between Semantic Prediction and Dialect Heard was found, which indicated that there were fewer looks to the target image for sentences containing neutral verbs when hearing the unfamiliar dialect (AAE) compared to the familiar dialect (GAE). A significant three-way among Semantic Prediction, Dialect Density, and Dialect Heard was observed, suggesting that as dialect density decreases, the log odds of looking to the target in neutral verb contexts also decreases. In other words, differences between looking at the target image in neutral contexts when hearing AAE and GAE is greater for children who produce fewer features of AAE. This 3-way interaction is visualized below in the top and bottom panels of Figure 6. The bottom panel shows a median split of dialect density for purposes of illustration and it can be observed that the listeners with high Dialect Density show a difference between the two Dialects Heard in the neutral condition, while the listeners with low Dialect Density do not. We also found a significant 3-way interaction among Semantic Condition, Dialect Heard, and Vocabulary indicating that holding dialect density constant, children with larger vocabularies as compared to children with smaller vocabularies leverage lexical-semantic cues more when hearing AAE. This 3-way interaction is visualized below in Figure 7. Similar to the other three-way interaction, it can be observed that the high-vocabulary group shows a larger differentiation between neutral and predictive Semantic

Conditions for the AAE stimuli, relative to the low-vocabulary group. All other main effects and interactions were not significant.

Table 9.
Summary of parameter estimates for the mixed-effects logistic regression model fit to the gaze data. These models were fit using maximium likelihood. Vocabulary is measured using the Peabody Picture Vocabulary Test-4th Edition using the Growth Scale Values (GSV) which are age normalized.

R Syntax: cbind(LooksTarg, LooksNotTarg) ~ Semantic Condition + Dialect Density + Age + Dialect Heard + Race + Vocabulary + Semantic Condition*Dialect Heard + Semantic Condition*Vocabulary + Dialect Heard*Vocabulary + Semantic Condition*Dialect Heard *Vocabulary + Dialect Density*Dialect Heard + Semantic Condition*Dialect Density *Dialect Heard + (Semantic Condition | Item) + (Dialect Heard + Semantic Condition| Participant)

Fixed Effects

	b	**SE**	**z**
Constant	.59	.13	4.55 ***
Semantic Condition	-.93	.14	-6.72***
Dialect Density (*DELV-ST*)	-.17	.16	-1.06
Age (*In Months*)	.23	.09	2.42***
Dialect Heard	-.05	.08	-.57
Race	.08	.24	0.35
Vocabulary (*PPVT-4 GSV*)	-.07	.14	.60
Semantic Condition * Dialect Heard	-.19	.01	-17.20***
Semantic Condition * Dialect Density	.10	.01	10.74***
Semantic Condition * Vocabulary	.01	.01	1.51
Dialect Heard * Vocabulary	.13	.10	1.32
Dialect Heard * Dialect Density	.09	.13	.45
Semantic Condition *Dialect Heard * Vocabulary	-.11	.01	-7.23***
Semantic Condition *Dialect Density * Dialect Heard	.08	.02	4.27***

Random Effects

Parameter	**Variance**	**SD**	**Correlations**
Participant-level Intercept	.22	.46	
Participant slope: Dialect Heard	.17	.41	-.52
Item-level Intercept	.18	.43	
Item slope: Semantic Condition	.24	.49	-.59

*p <.001***, p < .01 **, p < .05* , p ≥ .05 +*

Figure 6. The three-way interaction among Semantic Predictability, Dialect Heard, and Dialect Density. *Top panel:* Looks to target over time as a function of Semantic Prediction, Dialect Heard, and Dialect Density (median split). The purple window indicates the region of analysis identified by the cluster-permutation test. *Bottom panel:* Mean looks to target in that same window as a function of Semantic Prediction, Dialect Heard, and Dialect Density (median split).

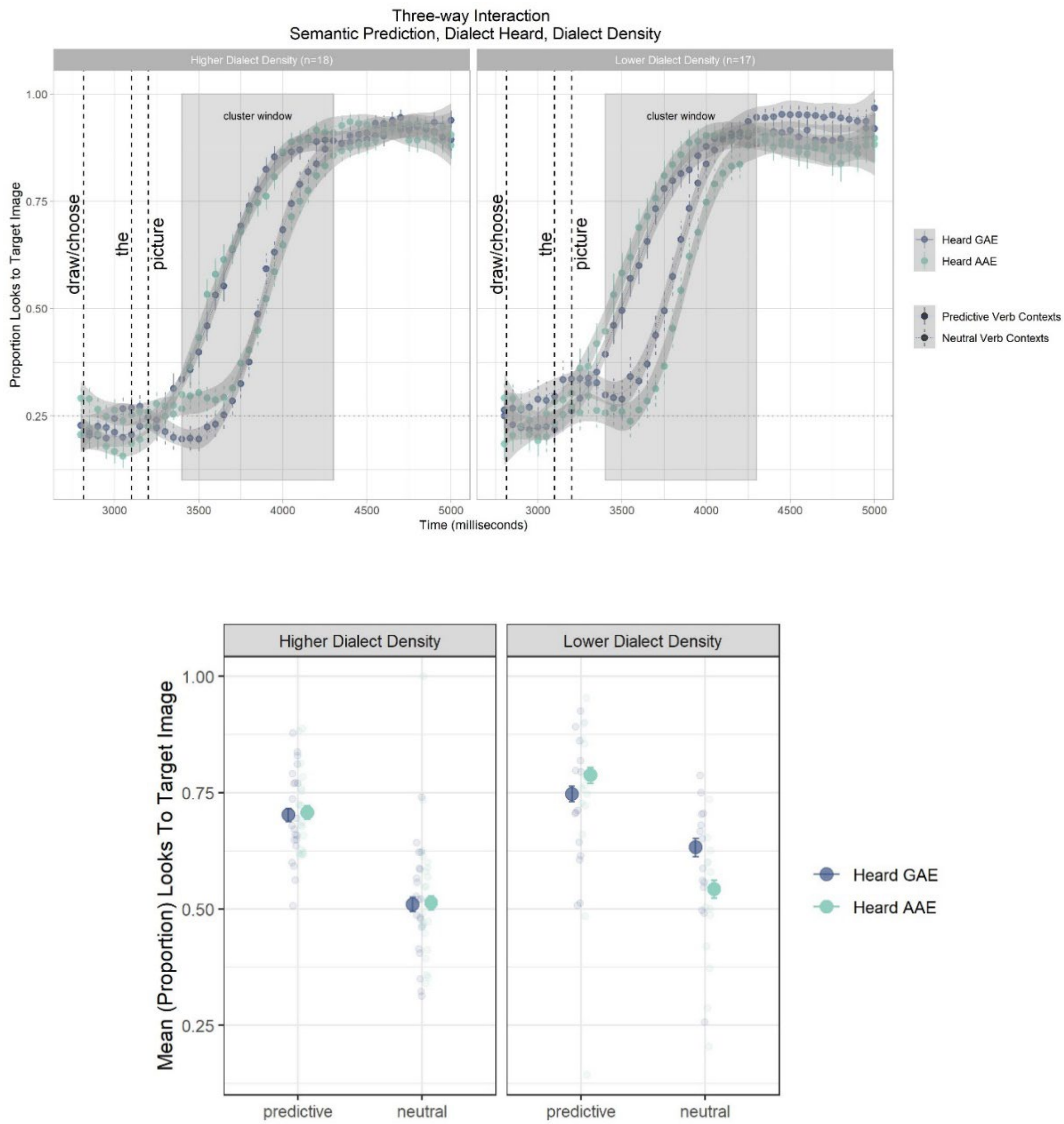
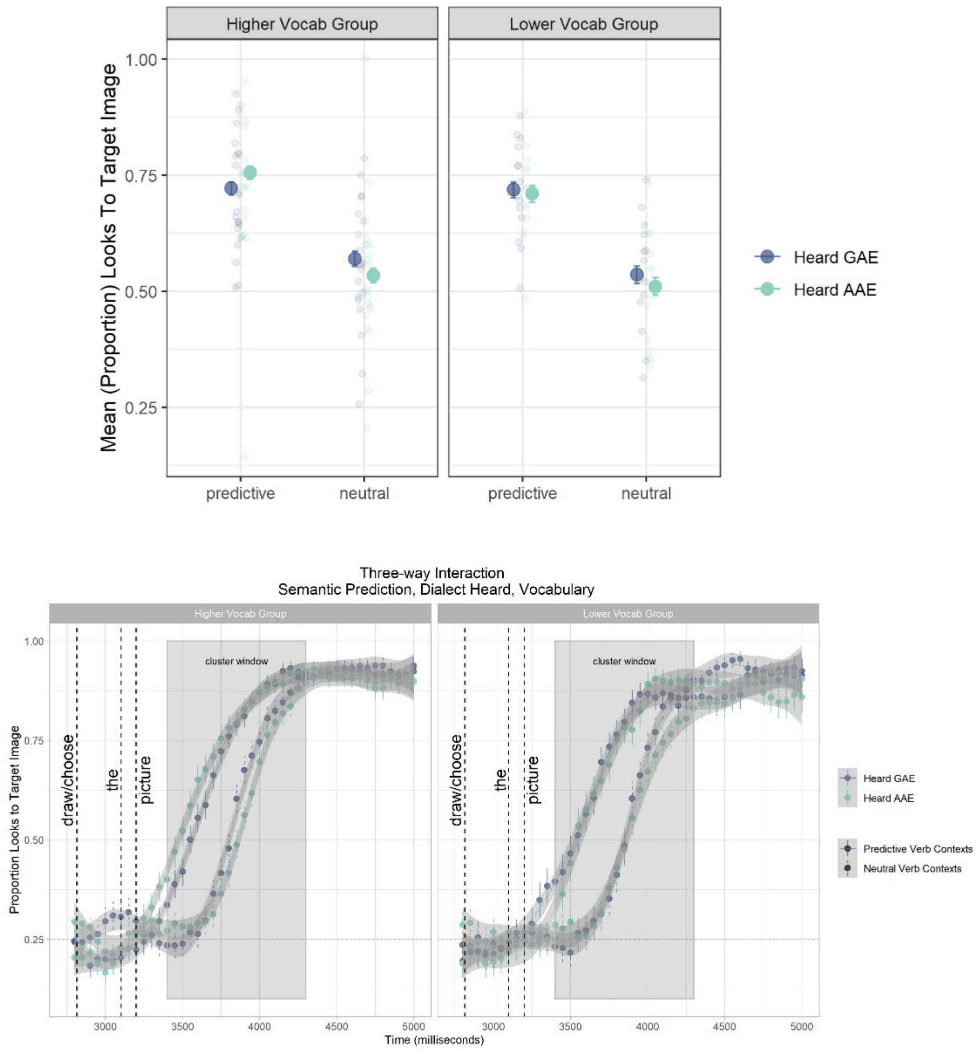


Figure 7. The three-way interaction among Semantic Predictability, Dialect Heard, and Vocabulary. Density. *Top panel:* Looks to target over time as a function of Semantic Prediction, Dialect Heard, and Vocabulary (median split). The purple window indicates the region of analysis identified by the cluster-permutation test. *Bottom panel:* Mean looks to target in that same window as a function of Semantic Prediction, Dialect Heard, and Vocabulary (median split).



Discussion

To summarize, Experiment 2 was designed to evaluate the effect of dialect variation on semantic prediction during sentence processing. The central finding of this study suggests that spoken language processing is influenced by dialect variation, however, these effects are restricted to contexts that provide limited facilitating semantic information (i.e., neutral sentences without predictive verbs). Further, our results suggest that differences in recognizing familiar words across dialects decrease among children with greater dialect density. Therefore, the results of this study provide some support for the perceptual analysis hypothesis, which proposes that dialect differences create challenges in mapping variable speech input to existing representations. The results indicate that dialect variation affects spoken language processing most when there is an absence of supportive higher-level cues (e.g., semantic context). It is plausible that neutral sentential contexts create more opportunities to attend to the signal. That is, in contexts where supportive linguistic information is reduced, the salience of dialect differences may be increased allowing dialect variation to have a greater impact on linguistic processing. This result aligns with previously advanced hypotheses regarding the function of supportive sentence contexts (Clopper, 2008; Creel, 2012). These theoretical frameworks posit that anticipatory information, across linguistic levels, supports the listener's comprehension of linguistic information in less optimal listening environments (e.g., non-native accented speech or speech in noisy listening contexts).

Additionally, our analysis included children with a broad range of dialect density scores (quantified as the proportion of NMAE responses on the DELV). We found that the relationship between semantic prediction and dialect differences was further moderated by dialect density (i.e., number of AAE features used in production). Specifically, dialect variation had a more pronounced influence on spoken language comprehension across children who produced more GAE-aligning responses on the DELV-ST (i.e., had lower dialect density). It is plausible that children with low dialect density in this sample were the least familiar with AAE grammar, regardless of their race. Thus, in contexts where there were fewer supportive contextual cues to guide word recognition, these children experienced more processing difficulty in the less familiar dialect (AAE). In contrast, children who produced a greater number of NMAE features on the DELV-ST (i.e., had higher dialect density) showed equal proficiency in recognizing words across dialect conditions. In this study, all participants were in elementary school (mean age was 7;9 and age range was 6;0 to 11;9) and had been exposed to GAE from their teachers. Perhaps this exposure to GAE in the classroom was enough to make both dialects familiar enough to not affect performance, at least for this task.

Finally, we address the additional findings involving vocabulary and age. In this study, there was a significant three-way interaction between semantic prediction, receptive vocabulary, and the dialects heard. There is a strong negative correlation between dialect density and receptive vocabulary size. Children with larger vocabularies had lower dialect density and were more likely to be speakers of GAE. This is the group that was identified by the three-way interaction among semantic prediction, dialect heard, and dialect density as showing poor word recognition in the neutral condition for the less familiar dialect (AAE). Alternatively, this interaction might indicate that children with larger vocabularies were better at leveraging lexical-semantic information provided by predictive verbs when hearing AAE. This finding provides meaningful opportunities for future research to further examine relationships among dialect use, vocabulary, and linguistic processing.

Interestingly, we did not observe a main effect of vocabulary. This is likely because our study included highly familiar words that on average had an age of acquisition close to 4 years old. Children in our study were older and most likely knew the target referents very well. It is also plausible that effects of vocabulary were masked by including age as a predictor because age and vocabulary are usually correlated. However, this is an unlikely explanation because excluding age as a predictor did not change the influence of vocabulary (see Table A3). In fact, the best fitting model for this study included main effects of both age and vocabulary. The findings with age in this study demonstrate that when controlling for potential variability in vocabulary size, older children demonstrate quicker and more reliable fixations to highly familiar words across semantic and dialect conditions.

General Discussion

To summarize, this paper presents two studies examining the effect of dialect variation on spoken language processing in two linguistic contexts: spoken word recognition in simple phrases (Experiment 1) and semantic prediction during spoken word recognition (Experiment 2). Together, these studies provide evidence that dialect variation does impact spoken language processing, but the effects are limited to contexts in which dialect variation is maximized (i.e., there are phonological and morphosyntactic differences) and semantically facilitating cues are greatly reduced (i.e., neutral sentential contexts). We interpret these finding in two possible ways.

One interpretation suggests that effects of dialect variation on spoken language processing are indeed explained by *perceptual analysis*; it can be more challenging to map an incoming signal to an existing representation in an unfamiliar dialect, even for words that are highly familiar to listeners. However, it would seem that these challenges are restricted to communicative contexts that provide suboptimal listening environments (i.e., those devoid of contextual or otherwise facilitating cues). If this is the case, then we must consider whether this framework is likely to account for the demands and challenges of spoken language comprehension that are observed in real world contexts. Furthermore, children with greater dialect density (AAE speakers) performed equally well in both dialects, suggesting that only a few years of school experience is adequate to mitigate the perceptual processing demands of the unfamiliar dialect.

These results suggest that the perceptual analysis hypothesis, at least on its own, may not adequately explain processing demands in an unfamiliar dialect and that experimental designs that utilize auditory stimuli divorced from the larger linguistic and non-linguistic context may not be adequate to examine the effect of dialect mismatch on spoken language comprehension. Spoken language comprehension is a result of complex interactions among the speaker's and listener's linguistic knowledge, subconscious or conscious linguistic choices, communication goals, audience design, and the environment itself (e.g., formal vs. informal situational contexts, etc.) (Babel & Mellesmoen, 2019; Bradlow & Bent, 2008; Walker & Campbell-Kibler, 2015). Several studies have already shown that listeners are sensitive to various factors both in terms of production (Arnold et al., 2007; Ratner, 2013; Ferreira, 2019) and comprehension (Babel & Russell, 2015; Carminati & Knoerferle, 2013; Kinzler, Dupoux, & Spelke, 2007; Kinzler, Shutts, DeJesus, et al., 2009; Münster & Knoeferle, 2018). More recently, social psychology has shown evidence that comprehension of speech is influenced by cues about the speaker's identity (Babel & Russell, 2015; Babel & Mellesmoen, 2019; Walker & Campbell-Kibler, 2015). These results suggest that the listener's comprehension of speech variability due to linguistic differences will necessarily depend on the listener's linguistic knowledge, understanding of the speaker's intent, and individualized social experiences with language variation and the individuals from communities who produce specific linguistic patterns. Given the dynamic and complex nature of most conversational contexts, future psycholinguistic experiments and models of spoken language processing, more generally, should be able to capture processing differences that arise as a function of linguistic differences across dialects and speaker-specific variables.

Chapter 3: Determining effects of socially enriched dialect contexts on semantic prediction

Experiment 3

A long-standing concern in education has been the persistent gap in achievement between African American and European American students, a gap that has recently widened because of the pandemic and protracted periods of remote schooling (NAEP, 2022). Decades of prior research have indicated that dialect variation is likely one of several factors contributing to achievement differences because many African American students speak African American English, a language variety that differs phonologically and morphosyntactically from General American English, the language of instruction in the United States. Empirical research on dialect variation and literacy achievement has demonstrated that linguistic differences between dialects make it more difficult to learn to read (Bühler et al., 2018; Charity et al., 2004; Gatlin & Wanzek, 2015; Washington et al., 2018, *inter alia*) and lately, more difficult to comprehend spoken language (Byrd et al., 2022, Edwards et al., 2014; Erskine, 2022a; Johnson, 2005; de Villiers et al., 2007; JM Terry, Hendrick, Evangelou, et al., 2010; JM Terry, Thomas, Jackson, et al., 2022). The literature on reading achievement has explained correlations between nonmainstream dialect use and literacy achievement in terms of the *linguistic interference hypothesis*, which asserts that difficulties learning to read are due to speech-to-print mismatches. It posits that differences between dialects create additional challenges for students as they try to match mental representations of speech to written language forms (i.e., processes important for decoding and spelling). The linguistic interference hypothesis has been supported by several studies that have shown that children who use more AAE features and have difficulty dialect-shifting have lower reading scores across the several literacy indices (i.e., standardized measures of decoding, spelling, and reading comprehension) (Gatlin & Wanzek, 2015; NP Terry, Connor, Petscher, et al., 2012; NP Terry, Connor, Johnson, et al., 2016; Washington et al., 2018; Puranik et al., 2020).

In contrast to developments in literacy research, studies on spoken language comprehension and dialect variation have primarily focused on establishing correlational evidence, but have not advanced clear, causal explanations that links dialect variation to spoken language comprehension. For example, Edwards et al. (2014) examined the comprehension of words that were lexically ambiguous or unambiguous among four- to eight-year-old AAE-speaking children. Words like *gold* would be considered ambiguous as they could be perceived as either *goal* [goʊl] or *gold* [goʊld] among AAE speakers. Edwards and colleagues found that AAE-speaking children were overall less accurate in comprehending words that were lexically ambiguous relative to unambiguous words that did not differ between AAE and GAE (e.g., *bus*). Earlier studies on the comprehension of third person singular -s show similar findings. Johnson (2005) and de Villiers and Johnson (2007) both tested children ranging from age 3 to 7 on their comprehension of tense/aspect and number agreement that is cued by third person singular -s. Across these studies they observed that AAE-speaking children did not comprehend this feature, while GAE-speakers began to show sensitivity to its grammatical interpretations approximately at age 5 to 6. These findings were later supported by Beyer and Hudson-Kam (2012) who assessed first and second grade AAE-speaking children on their comprehension of a variety of GAE morphological markers. Beyer et al. found that while GAE-speaking children demonstrated comprehension of some tense markers (i.e., past tense -ed, future contractable ‘ll, and third person -s), AAE-speaking children only showed accurate comprehension of morphological forms that overlapped between GAE and AAE (i.e., plural -s).

One explanation that might account for the findings above comes from the perceptual processing literature in psycholinguistics. Specifically, a perceptual processing framework asserts that differences between AAE and GAE will make it harder to map a rapidly unfolding auditory signal to the listener's mental representations, the ***perceptual analysis hypothesis.*** Similar to the linguistic interference hypothesis, it too presents dialect as a mapping problem, but instead of focusing on the role of speech-to-print differences, it emphasizes the relevance of the auditory signal. From a purely linguistic perspective, this is a plausible assertion because word recognition and spoken language comprehension more generally require listeners to encode auditory information and retrieve relevant meanings. Additionally, research on accent variation has shown consistent evidence that variability in the acoustic-phonetic realization of words disrupts processing. Since similar disruptions might be observed when listeners hear different dialects, a perceptual analysis framework may be helpful for characterizing comprehension differences. Moreover, it provides a clearer causal link between dialect use and spoken language processing because it creates clearer predictions about the way disruptions arise (i.e., ruling out competing lexical candidates due to linguistic ambiguity or disrupting the completion of academic tasks because of depleted cognitive resources).

To investigate the validity of the perceptual processing account, Erskine (2022a) examined the effects of dialect differences between AAE and GAE in two real-time spoken language processing experiments: a) spoken word recognition in simple phrases and b) use of semantic prediction during word recognition. In an eye-tracking study, children (5-8 years old) who varied in their use of AAE and GAE heard words in simple phrases or sentences and their gaze patterns to one of four objects were monitored as they listened and selected images on the screen. Erskine et al. found that children had less efficient word recognition in longer sentences that contained few facilitating lexical-semantic cues (i.e., sentences that contained neutral verbs, *find,* as compared to predictive verbs, *eat*). Moreover, this result was observed only for children who had very little experience with the less familiar dialect. Their results provide preliminary evidence that partially supports a perceptual analysis explanation. However, because the effects of dialect were small and limited to sentences with neutral verbs, the degree to which the perceptual processing hypothesis fully explains relationships between dialect variation and spoken language comprehension is still an open question. One central limitation of the studies described above is that the authors examined dialect variation in very controlled and arguably unrealistic listening environments. Specifically, they studied children's comprehension of dialect differences in the absence of additional information about the speaker's identity (i.e., socioindexical properties of the speaker). While this methodological approach aligns with a perceptual analysis account, which emphasizes the salience of the signal, it is limited in its ability to capture social nuances that are inherent to dialect use. Additionally, it provides a very narrow understanding about how spoken language comprehension is further modulated by socioindexical characteristics of a speaker.

Acknowledging and addressing this limitation is important for a few reasons. First, dialects are usually nested in linguistic communities that differ systematically by any number of social factors. AAE is a prime example of this kind of systematic variation because it is a raciolect that is associated with a specific group of speakers (i.e., spoken by many African Americans). Additionally, children's early experiences with hearing AAE and GAE are associated with their own lived experiences with particular groups of speakers that produce these varieties. Second, in any communicative context, listeners encode information about the speaker that goes beyond the linguistic signal and includes salient non-linguistic properties of the signal (i.e., the speaker's age, sex, and race/ethnicity).

A growing segment of the speech processing literature has demonstrated that spoken language comprehension is influenced by listeners' awareness of nonlinguistic aspects of the speaker's identity, either through *implied* (not visually salient) or *overt* (i.e., visually salient) speaker identity cues. Social characteristics of speakers such as their age (Drager, 2011; Harrington et al., 2007), geographic location (Hay et al., 2006), socioeconomic status (Hay & Drager, 2007; Lawrence, 2017), sex and/or gender (Munson, 2011; Winn et al., 2013; Munson & Logerquist, 2017), and race/ethnicity (Babel & Russell, 2015; Babel, 2022; Staum Casasanto, 2008) have been shown to affect speech perception. In this next section, we briefly review evidence from studies that have focused on visual aspects that index race and ethnicity, as the central objective of this paper is to understand how listeners comprehend socially-marked language varieties that differ across individuals of diverse racial backgrounds.

A number of studies have shown that pairing auditory stimuli and visual information in the form of faces impacts spoken language comprehension. For example, Babel and Russell (2015) demonstrated that European Canadian adult listeners understand American English in noise with greater accuracy when paired with a photograph of a European American face, but with poorer accuracy when paired with a photograph of a Chinese Canadian face (see also Kang & Rubin, 2009). Other studies have found that European Americans' comprehension of Mandarin-accented speech in quiet improves when the signal is paired with photographs of Asian speakers (McGowan, 2015). Staum Casasanto (2008) found that listeners were more likely to anticipate zero-marking of final consonants (i.e., /t/) in words (e.g., /mæs/ instead of /mæst) when the auditory stimuli were paired with the face of an African American. Additionally, research with multilingual communities has shown that sensitivity to linguistic and non-linguistic sources of information about the speaker's identity impacts spoken language comprehension. For example, Kutlu, Tiv, Wulff, and Titone (2022) observed effects of race and perceptual ratings of accent on transcription accuracy among multilingual participants across two locales that differed in bilingual experiences and tolerance. Specifically, they found that for locales where bilingualism was less normalized, social cues (i.e., racial background of speakers) played a prominent role during speech perception. Transcription accuracy was greater for auditory stimuli paired with photographs of individuals with European descent relative to photographs of individuals of Southeast Asian descent. By contrast, this same effect was not observed in locales where bilingualism was more normalized.

In summary, there is evidence across a range of studies with different methodologies that spoken language comprehension is altered by the presence of nonlinguistic *socioindexical* speaker characteristics. However, the majority of this work has focused on adults. An important question that remains is whether children demonstrate sensitivity to visual cues to race and ethnicity and whether this information impacts linguistic processing. A few studies have shown that children are sensitive to racial cues early in life, even as early as 3 months of age (Bar-Haim et al., 2006; Kelly et al., 2005; Wheeler et al., 2011). For example, some studies have shown that at 9 months of age children demonstrate a preference for faces that are familiar or similar to the child's race/ethnicity (Anzures et al., 2009; Kelly et al., 2009; Wheeler et al., 2011) relative to faces that are different in race. More importantly, research has also shown that children's social preference and spoken language processing can be influenced by visual (i.e., race/ethnicity) and auditory (i.e., accent/dialect) speaker cues.

Kinzler et al. (2009) examined 5- and 6-year old children's selection of similar-aged peers that differed in accent and race as friends. Children were presented with photographs of either European American or African American children. For each visual condition, half of the auditory stimuli contained English words produced in a French accent and the remaining half were English words produced in a General American English accent. They found robust evidence that when auditory information is absent, children showed a preference for choosing to be friends with a European American. However, when auditory and visual information was present, children consistently chose to be friends with speakers of a GAE accent regardless of race. Weatherhead and colleagues (2018, 2021) conducted research examining the effects of speaker race on real-time spoken word recognition and on children's adherence to mutual exclusivity during word learning. In each study, they found evidence that across language processing tasks (i.e., comprehension or learning), visual aspects of the speaker's identity influenced their performance across different speaker condition. Specifically, when words were presented by a speaker of the same racial background as the child, 16-month-olds showed a clear processing advantage for recognizing words in a familiar accent as compared to an unfamiliar accent. However, when words were presented by a speaker of a different race, children looked equally to target images across familiar and unfamiliar accents (Weatherhead & White, 2018). In a more recent study, Weatherhead et al. (2021) demonstrated evidence of word learning flexibility in the presence of visually salient socioindexical speaker cues (i.e., race). First, they showed that when monolingual or bilingual children learn words by speakers of the same, each linguistic community demonstrated evidence of mutual exclusivity. This contrasts with some trends in the literature which have suggested that bilingual children do not do this to the same degree. Second, for conditions in which the speaker was of a different race, monolingual children relaxed their adherence to mutual exclusivity principles when learning new words.

Overall, these results suggest that children utilize socioindexical properties of the speaker to make generalizations about the speaker's use of language and communication intent in ways that vary across a variety of tasks (e.g., word recognition or word learning). These results also suggest that children track relevant speaker cues in ways that influence their ability to reason about speakers, which in turn guide their processing strategies and efficiency. When properties of the speaker are familiar or similar to children, they may develop strong expectations about what people will say and the meanings that are conveyed. However, when properties of the speaker are less familiar, it is possible that their ability to reason about the speaker is attenuated and this can impact spoken language processing in complex ways that either facilitate or inhibit spoken language comprehension. Moreover, we can anticipate facilitating effects of speaker identity when socioindexical information supports the listener in reasoning about and creating accurate model of speakers' mental and/or knowledge states. Additionally, experience with different communities of speakers and knowledge of their identity provide listeners with assumptions that either facilitate or impede them in building linguistic representations of the incoming signal.

This interplay between linguistic and nonlinguistic aspects of the speaker's identity and the relevance of visual socioindexical cues for processing dialect variation is not captured by perceptual analysis explanations. Whereas the perceptual analysis hypothesis predicts that processing of dialect differences will be modulated by linguistic information (e.g., predictive verbs or context provided through conversation topics), it makes no predictions about how the analysis of the signal will be affected by non-linguistic properties of the speaker. This information is important for identifying causal mechanisms for the effect of dialect differences on literacy achievement because metrics of academic achievement are an assessment of information that the child has learned by encoding information through spoken instruction and learning opportunities by an individual (i.e., educators and even classmates). For most children, information from the auditory signal and the speaker's identity will interact in ways that meaningfully influence spoken language processing (Bricker & Pruzansky, 1966; Creel & Bregman, 2011; Ladefoged & Broadbent, 1957).

Experiments 3.1 and 3.2 are designed to test this claim more directly. Each of these experiments examine school-age children's (6- to 11-years-old) integration of visual and linguistic information in a spoken language comprehension task (i.e., spoken word recognition in semantically facilitating sentences). Moreover, we do so in two communities of listeners, European American children with relatively limited experiences with stigmatized dialects of English (i.e., AAE) and African American children with more variable experiences with speakers of dominant and stigmatized dialects of English. Our focus is limited to these two specific communities because research on the achievement gap has typically focused on differences between African American and European American students. Additionally, we examine this question in a group of children who are older than have been previously studied (Weatherhead & White, 2018; Weatherhead et al., 2021) because older children have more variable experiences with different groups of speakers than 16- to 18-month-olds. Lastly, we use a more demanding task than previous studies because in naturalistic communication contexts children frequently hear words in longer utterances and differences between AAE and GAE are more perceptually salient in sentences as compared to single words.

To examine children's integration of visual and linguistic aspects of the speaker's identity during spoken language comprehension, we examine three different permutations of speaker group conditions: a) hearing a dialect that matches the listener's primary dialect and seeing the face of someone of same racial background b) hearing a dialect that does not match the listener's primary dialect and seeing the face of someone of a different racial background, and c) a condition in which either the dialect or face of the speaker matches the listener's. Given the dynamic and complex nature of how identity intersects with social experiences among listeners, we anticipate that different listener communities might prioritize socioindexical cues to speaker identity differently. This assumption reflects the notion that the cues that are most informative for reasoning about a speaker will depend on the listener's lived experiences with dialect variation and identities that exist within a certain community of speakers.

For the remainder of this paper, we report results for each listener community separately. We take this approach because our administration procedures differed across groups due to challenges associated with the pandemic (further details provided in the Methods section for African American listeners). Experiment 3.1 describes the experimental procedures and findings for European American listeners and experiment 3.2 discuss differences in experimental procedures and the findings for African American listeners.

Experiment 3.1: European American Listeners

How will speaker identity cues influence European American children's recognition of words in facilitating contexts? The evidence provided by prior studies is mixed. On one hand, there is evidence that children's comprehension may be based on their familiarity and/or experience with different speakers (Adank et al. 2009; Brungart et al., 2020; Creel & Bregman, 2011; Nygaard et al., 1994). For example, as described earlier, Weatherhead et al. (2018) found that for speakers of the same race, listeners demonstrated an accent familiarity preference. However, in that same study, Weatherhead et al. also showed that comprehension may be less impacted by accent differences if listeners are less familiar with a person's race (i.e., children demonstrated relaxed linguistic expectations which actually facilitated word recognition). This indicates that children are potentially flexible in their recruitment of speaker identity cues depending on the task (i.e., simple word recognition) and speaker context (i.e., familiar or unfamiliar race). However, it is not clear whether results observed with infants will scale up to school-age children or more complex linguistic tasks. An additional consideration is how accent and race cues interact with other social processes (i.e., selecting friends or selecting credible learning partners). While a large number of studies dated back to the 1950s (Stevenson & Stewart, 1958; Fox et al. 1973, *inter alia*) have found that European American children will select images of European American children over African American children as potential friends, Kinzler et al. (2009) found that European American children selected images of African American children paired with a more mainstream American-accented English over European American children paired with French-accented English as potential friends.

There has not been a lot of work on the effect of speaker identity on spoken language comprehension among school age children. Therefore, we base our predictions on previous findings from Weatherhead et al. (2018) because the variables of interest in their study (race and accent) are also of interest in this study but instead of accent, we look at dialect differences between AAE and GAE. Additionally, Weatherhead et al. examined these factors in a task that is a simpler variant of the task used in this experiment (i.e., spoken word recognition in semantically facilitating contexts). We also take into account the findings from Kinzler et al.

(2009) which shows that school-age children prioritize accent cues over race. Therefore, this study evaluates two different predictions:

1. It is possible that **children's word recognition will be guided by dialect-related cues over and above race.** Since older children have more variable experiences with speakers, we may observe findings that are similar to Kinzler et al. (2009) in which listeners demonstrate more efficient and accurate word recognition of speakers of the same dialect as the primary dialect of the child, regardless of the speaker's racial background. Additionally, we predict that children will demonstrate poorer word recognition for speakers of a dialect that differs from their primary dialect.
2. Alternatively, it is also possible that **word recognition will not differ across speaker conditions.** If children perform similar to the infants in Weatherhead et al. (2018), we predict that children will demonstrate an effect of dialect variation for speakers of the same race, but not for speakers that differ in race. In context of the experimental conditions we included, this means that children will demonstrate efficient word recognition for speakers in which they see a face that is the Same Race and Hear GAE (i.e., see a photograph of a European American and hear GAE). Second, for speakers of a different race, children will demonstrate minimal differences between conditions in which they see a face that is a Different Race and Hear AAE or GAE (i.e., seeing a photograph of an African American and hearing AAE and seeing a photograph of an African American and hearing GAE, respectively). In other words, differences across all three conditions will not be observed.

Using the visual world paradigm, we examined how visual and auditory cues to speaker identity combine to influence spoken word recognition efficiency in sentences with a facilitating semantic context.

Methods

Participants

Listeners

Twenty-three European American children (12 female) between the ages of 72 and 138 months (mean=105, sd=18) were included in this study. Children from areas near Washington, D.C., Maryland and Virginia (the same region as in Experiment 2) were invited to participate during the March-to-August 2022 period of the pandemic. An additional 8 children were recruited but not included in the final sample for the following reasons: experimenter error or equipment failure (n = 5) or missing data that exceeded the threshold for inclusion (n = 3).

Children were primarily recruited from a University-based consortium of families who expressed interest in participating in child development research. Families from this database were provided information about the study and invited to participate through email. Other families who participated heard about this study through word-of-mouth from families who had previously participated; these children usually attended the same school as previous participants or lived in the same neighborhood (n=5). Families filled out a demographic survey that asked about the highest level of maternal education; the majority of participants were from families with higher levels of maternal education (e.g., received a college degree or more). All children completed in-person research visits at the University.

Table 1
Demographic summary of the 23 participants and their scores across standardized assessments administered in Experiment 3.1.

Female	12
Mean Age (in months)	105 (18) *range=72-138*
NIH PVT Toolbox *Mean Raw Score* (SD)	83 (8) *range=67-95*
Mean Standard Score (SD)	113 (10) *range=91-131*
Maternal Education	
Low (GED or high school diploma)	1
High (college degree or more)	21
Declined to report	1
DELV-Screening Test (Part I)	
Speaker of GAE	19
Speaker with some variation[3]	4

Auditory Stimuli

Sentences and nouns. The experimental stimuli in this experiment consisted of 54 sentences. Of the sentences, 27 were the same test sentences as in Erskine (2022a). These 27 sentences contained a predictive verb (i.e., *Every day, Caleb helps his friend **read** the book*). Nine sentences from Erskine (2022a) that contained a neutral verb (*Every day, Caleb helps his friend **choose** the book*) were also included to confirm that there was a facilitating effect of the predictive verb. Similar to Erskine (2022a), the neutral verbs were matched to predictive verbs. Six additional predictive verb sentences were adapted from Blomquist et al. (2021) to increase the number of items in the study. Further, an additional 18 sentences were included as filler items

[3] As described in Language Measures section, children were administered the Diagnostic Evaluation of Language Variation (Seymour et al. 2003) to confirm their primary dialect. Four children received a criterion score that categorized them as speakers with “some variation.” There are two reasons why children might have received this score: a) they were speakers whose dialect include a small number (1-2) of NMAE features or b) they were children at risk for developmental language disorders (DLD). However, all children received standard PVT scores that were within normal limits. Because the range of dialect scores (calculated as proportion of items containing an NMAE feature relative to the total number of NMAE and GAE responses on the DELV-ST) was very limited (range = 0-.27, mean = .08, sd = .09), we did not include a measure of dialect density in the statistical analysis.

to reference a non-target referent in the study. Filler trials contained neutral verbs that were not matched to a predictive verb; these items were included to increase children's visual attention to the other images within the 2x2 display. A complete list of the experimental trials can be found in Table G1. All referents in the study were monosyllabic and bisyllabic nouns that had an age of acquisition that was less than 7 years (Kuperman et al., 2012). All words were pictured with images or photographs available from public online databases.

Speaker recordings and auditory stimuli processing. Sentences were recorded by three female 21- to 25-year-olds who self-identified as belonging to one of the following ethnicity and linguistic communities: a) a European American monodialectal speaker of GAE, b) an African American monodialectal speaker of GAE, and c) an African American bidialectal speaker of AAE and GAE.

All auditory stimuli were recorded using a Shure SM51 microphone in a sound-attenuated booth. Sentences were later normalized for duration across all speakers across two segments: from the onset of the verb (*bake*) to the determiner (*the*) and from the onset of the noun (*cake*) to the end of the noun. Lastly, all sentences were normalized to the sound level of 70dBSPL and 1000 milliseconds of silence was added to the beginning and end of the sentence. This resulted in sentences that were 6,023 milliseconds in length. The critical verb (*bake*) was time-locked at the 3,158 milliseconds, the determiner (*the*) began at 3,460 milliseconds and the target noun onset (*cake*) was time-locked to 3,568 milliseconds.

Visual Stimuli

Object images

All words were pictured with images or photographs available from public online databases. Four images of each referent were selected because children would see a referent across multiple trials. We used images that were tested in Erskine (2022a) and Blomquist et al. (2021). Images across these studies were normed in preschool and kindergarten classes at the Center for Young Children, a school for children that is affiliated with the University (see Erskine [2022a] and Blomquist et al. [2021] for visual stimuli preparation and norming procedures).

Speaker images

To evaluate the contribution of the speaker identity cue, *race/ethnicity*, photographs of each "speaker" were included in this study. Photographs were taken from the Chicago Face Database (Ma et al., 2020), a repository of photographs of 597 unique individuals. This database contains faces of individuals who self-identified as males or females from the following race and ethnicity backgrounds: Asian, African American, LatinX, and European American. For additional details and norming procedures visit (Ma et al., 2020**)**. For the purposes of this experiment, we initially selected eight images, four unique pictures of European American or African American females with happy, closed-mouth expressions to ensure an implied friendly demeanor. The advantage of using this database is that photographs of each individual exhibited uniform poses, attire, and expression. In other words, all individuals wore the same item of clothing, had similar facial expressions (e.g., a close-mouthed smile) and posed facing forward.

A growing body of research (Zebrowitz & Montepare, 2008) has shown that less desirable social attributes are typically associated with individuals from minoritized backgrounds on the basis of visual cues. To minimize the extent to which socially-biased inferences would interfere with the experimental task, an informal panel of adults from diverse cultural and linguistic backgrounds was convened to select images that activated more neutral assumptions about the speaker's social prestige. This process resulted in the final selection of 4 European American photographs and 3 African American photographs. All images were edited using a commercial picture editing software Adobe Photoshop (Adobe Inc., Version CC 2019). Speaker images were cropped to 500x500 pixels and superimposed on a canvas of 600x600 pixels with a grey background.

Finalized experiment lists

The final 54 sentences were distributed across three speaker group conditions: a) *Different Race|Heard AAE;* seeing an African American face and hearing AAE, b) *Different Race|Heard GAE;* seeing an African American face and hearing GAE and c) *Same Race|Heard GAE;* seeing a European American face and hearing GAE. In each condition, children heard 18 sentences: 9 contained predictive verbs, 3 contained neutral verbs, and 6 contained filler items. Four experimental lists were created based on the permutations of the selected images of Speakers across the Speaker Group conditions. Because the number of photographs of African American and European American was unbalanced, across each list, one of the four European American images was randomly selected without replacement and assigned to the condition in which listeners saw a European American photograph and heard GAE. Additionally, in each list, two of three African American photographs were pseudorandomized and each photograph was then assigned to one of the two remaining speaker group conditions. This permitted the creation of lists that contained a unique photograph-speaker group condition pairing. Additionally, the image that was randomly assigned to a specific speaker group condition remained fixed over the duration of the entire experiment. This was crucial to ensure that participants had consistent opportunities to track the visual and auditory information provided by each speaker. Table G2 in Appendix G contains a table illustrating an example of image-condition pairings across one of the experimental lists.

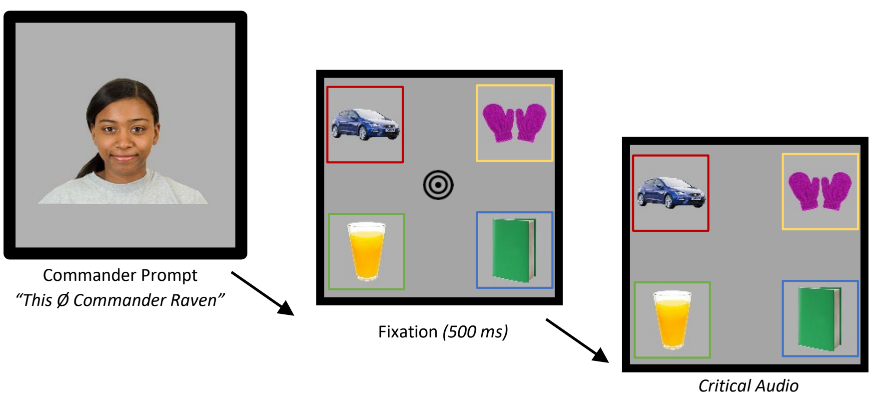
A within-subjects design was used. Sentences across all three speaker group conditions were intermixed. All items were pseudorandomized offline and then administered in a single block of eye-tracking. The trials were grouped into 6 smaller chunks for administration purposes, separated by a visual reward (i.e., a picture of missing "item" in their quest) to maintain children's participation throughout the task and to provide breaks when required.

Visual World Paradigm

Procedure. Gaze patterns were monitored using either an automatic SR Research Arm-mounted Eyelink 1000 Plus (n = 12) or a portable Duo Eye-tracking System (n = 11). In contrast to Experiment 2, this experiment was designed to model a space galaxy quest game in which children were rewarded with a missing "item" or "crew mate" while participating in the experimental task. At the start of the experiment, children were informed that they would be introduced to specific commanders who would tell them a sentence. Children were instructed to listen carefully and to select the image that best-matched the sentence that was heard using a touchscreen stylus. After the initial instructions, children were introduced to all three commanders one at a time, and were presented with an auditory introduction from each commander that was recorded in the dialect assigned to each speaker group condition (*"Hi. My name is Commander [name]. I'm going to tell you something about one of the pictures on the screen. Listen carefully!"*). This procedure was used to increase children's visual attention to the speaker images and auditory stimulus throughout the course of the experiment.

Trial event sequence. At the beginning of each trial, children were shown an image of one of the commanders while they heard a short prompt that reintroduced the Speaker (*"This is Commander [name]*." After 500ms, the image of the speaker disappeared and four images of the noun referent appeared on the computer monitor in a 2x2 matrix. The images remained on the screen for 900ms so that children could familiarize themselves with the images. Following this period, a fixation attention getting video (e.g, a bullseye) appeared in the center of the screen. Children's fixation to the attention getter for at least 500ms initiated the onset of the critical auditory stimulus. If the children required additional prompting, the experimenter would remind children to listen to the entire sentence and then select the picture that matched what they heard. The trial terminated after children selected an image using the stylus, which then initiated the start of the next trial.

Figure 1. Trial sequence of an example trial (*speaker group condition*: African American Speaker/Heard AAE).



Language measures

In addition to the visual world paradigm, children were administered two assessments, one that evaluated receptive vocabulary and one that evaluated children's use of U.S. English dialect patterns. The measure of receptive vocabulary was taken from the NIH Toolbox Cognition Battery which is appropriate for children between the ages of 3 and 85 years old.

Additional details of the development and assessment and scoring procedures can be found in (Weintraub et al., 2013). Participants were shown a 2x2 array of images on an iPAD and heard an audio recording of a word. Children were provided with instructions to select the image that best represented the meaning of the word that was heard. The receptive vocabulary measure uses an adaptive testing procedure in which the number of items administered depend on children's accuracy of the previous word. Children are administered items one at a time until the task is completed. This measure provides a variety of scores based on item response theory. In the statistical analysis, we used the "Uncorrected Standard Score." Although the term *standard* is in the label, it is a raw score that evaluates the participant's performance or vocabulary knowledge relative to other children in the United States. We use this term for consistency with the literature and for replicability purposes in future studies. This raw score provides an overall index of children's vocabulary knowledge. We also administered the DELV Screening Test Part I (Seymour et al. 2003) to categorize children's dialectal patterns and confirm their primary dialect.

Results

Below, we report the procedures and results for three analyses: a) response accuracy, b) the cluster analysis, and c) the final analysis examining the effect of speaker identity cues on word recognition in semantically facilitating sentences. An effect of semantic prediction was visually observed (see Appendix F). Therefore, as planned, the central analyses were restricted to the sentences with a predictive verb.

Response accuracy. Overall accuracy in this experiment was above 99% across all speaker group conditions (i.e., children demonstrated ceiling level performance). Across all of the participants, only two errors were observed. Moreover, these errors occurred in speaker group conditions in which listeners heard GAE from either a European American or African American speaker. No response errors were observed in the condition where listeners heard AAE produced by an African American. A generalized linear model with accuracy coded as binary (0,1) was used to examine whether accuracy was influenced by Speaker Group, Vocabulary, Age, and pairwise interactions between Speaker Group and each of the other aforementioned independent variables. Speaker Group was dummy coded such that the reference level was the condition in which the listener heard GAE and saw a photograph of a European American (i.e., *Same Race|Heard GAE*). All main and interaction effects were not statistically significant (see Table E1 in Appendix E).

Eye-tracking analyses and results. Data cleaning and screening occurred in two phases. First, a customized script (Github Script, Mahr, 2017) implemented a *deblinking* procedure which interpolated short windows of missing data, up to 150ms, if the child appeared to fixate on the same image before and after a *missing* window of data. The assumption of the interpolation process is that 150ms is too short a window for children to fixate to a different area of interest. The quality of the data was then examined in the time region of 3158 to 5568ms, the region beginning at the onset of the critical verb (read/choose) to 2000ms after the onset of the target referent (book). Children were excluded from all further analyses if they did not have an adequate amount of data in at least one block of eye-tracking. EyetrackingR evaluated trackloss across trials and removed trials with over 50% missing data.

The cleaning procedures resulted in 490 out of 675 trials with reliable data. Further analysis of the missing data revealed that the proportion of missing data did not differ by condition (i.e., Speaker Group), $F(2, 2897) = 2.57$, $p = .08$. All subsequent analyses were carried out only with trials with correct target responses.

Table 2.
Summary of the track loss and sample analysis before trial-exclusion. AA is the abbreviation for African American and EA is the abbreviation for European American.

Speaker Group	Number of Trials	Mean Samples	Percent Track Loss (%)
Saw AA Speaker: Heard GAE	225	2512	52
Saw AA Speaker: Heard AAE	225	2533	54.3
Saw EA Speaker: Heard GAE	225	2562	54.3

Table 3.
Summary of the track loss and sample analysis after trial-exclusion. AA is the abbreviation for African American and EA is the abbreviation for European American.

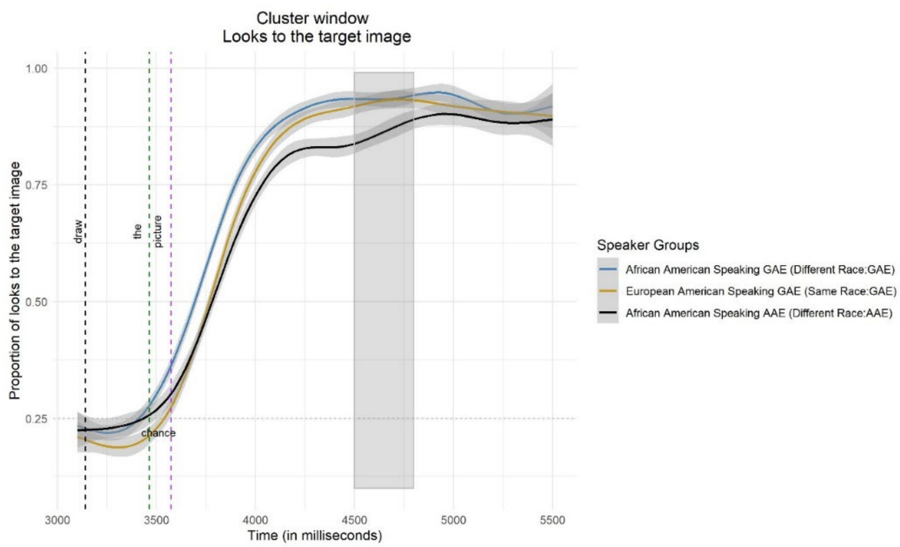
Speaker Group	Number of Trials	Mean Samples	Percent Track Loss (%)
Saw AA Speaker: Heard GAE	171	2387	45.7
Saw AA Speaker: Heard AAE	162	2385	44.8
Saw EA Speaker: Heard GAE	156	2442	46.0

Cluster analysis. As in Erskine (2022a), a cluster-based permutation analysis was used to identify the window of analysis. This method was chosen because the sentences are long and there are two points of facilitating information in each sentence (i.e., the verb and noun) and an objective method was needed to identify a window of analysis. Prior to conducting the cluster-based permutation tests, the data was restricted to a general window of time that began at 3100ms and ended at 6000ms. This window reflects the onset of the critical verb to 2400ms after the onset of the target referent. Additionally, this window was also informed by empirical evaluation of the data in which looks hovered at chance across all three conditions until the period after the noun started to unfold and evaluation of the previous study in which effects were observed in a later window than found in prior studies (Blomquist et al. 2021, Erskine et al. 2022a).

To isolate a region of time for subsequent analysis, we fit a mixed effects model predicting the empirical logit of fixating to the target image as a function of Speaker Group, a by-participant random intercept and slope for Speaker Group, and a by-item random intercept. The speaker group condition *Different Race|Heard GAE* (i.e., seeing an African American and hearing GAE) was excluded from the cluster analysis and included speaker group conditions that were either the most similar to participants (*Same Race|Heard GAE,* seeing a European American Speaker and hearing GAE) or dissimilar to participants (*Different Race|Heard AAE,* seeing an African American Speaker and hearing AAE). This approach was taken for several reasons. First, the permutes package conducts pairwise comparisons across several conditions introducing additional subjectivity in window selection when comparing regions across more than two levels of a given factor. Second, the Speaker Group condition, *Different Race|Heard GAE,* was excluded because it is not clear how to interpret differences that arose 68ms prior to the onset of the target referent between *Different Race|Heard GAE* condition and the other two conditions. A qualitative inspection of the audio recordings revealed that the speaker of the *Different Race|Heard GAE* speaker group condition (i.e., hearing GAE from an African American Speaker) produced sentences with a very clear and steady voice and a highly resonant voice quality. Moreover, many children in the study reported that they liked listening to the African American speaker producing GAE. Therefore, early arriving differences in this condition may reflect something unique about this speaker's voice such that many children preferred listening to this particular voice, independent of the experimental manipulation. Third, this study seeks to discuss performance among European American listeners in Experiment 3.1 to African American listeners in Experiment 3.2. This is not possible when regions of time are vastly different for each listener community (see Experiment 3.2 cluster analysis for reference).

The cluster-based permutation identified seven contiguous 50ms time bins from 4500ms to 4800ms (i.e., 932ms after the onset of the target word) in which the main effect of Speaker Group met the criterion threshold for significance. Figure 2 depicts the window of time identified.

Figure 2. Looks to target as a function of time and speaker group condition. The analysis window identified by the cluster analysis is highlighted in grey.



Generalized linear model analysis and results. To examine the effect of speaker identity cues on spoken language comprehension, we fit a logistic mixed effects model to the region identified by the cluster analysis (i.e., 4500 to 4800ms). The model included independent variables of Vocabulary (PVT uncorrected score which is equivalent to a raw score for other vocabulary measures such as the *Peabody Picture Vocabulary Test* [Dunn & Dunn, 2007; Dunn, 2019]), Age (in months), Speaker Group (all three conditions), and all pairwise interactions between Speaker Group and the other independent predictors. All continuous independent variables were centered and the categorical predictor Speaker Group was contrast coded using simple coding. The results of this approach are relatively similar to dummy coding in that each level is compared to a reference level. However, within this model, the intercept corresponds to the grand mean of the log odds of fixating on the target image for children of average age and vocabulary. In the fitted model, the first contrast of Speaker Group is interpreted as the difference in looking at the target image when seeing an African American Speaker and hearing GAE (*Different Race|Heard GAE*) from a reference level of seeing an EA Speaker and hearing GAE (*Same Race|Heard GAE*). The second contrast of Speaker Group evaluates the difference in looking at the target image when seeing an African American Speaker and hearing AAE (*Different Race|Heard AAE*) from the reference level. This allowed us to directly evaluate the hypothesis that children from EA backgrounds would be more likely to prioritize signal-related cues above visual cues indicating ethnicity of the speaker. Lastly, unlike previous models, we did not include maternal education as a predictor because children mostly came from families with higher levels of maternal education, only one family reported having maternal education levels that differed from a college degree (i.e., less than high school).

We found a main effect of the second contrast of Speaker Group. That is, children looked less at the target image when they saw an African American Speaker and heard AAE as compared to when they saw a photograph of a European American speaker and heard GAE. There was no significant difference between the other two Speaker Group conditions, *Different Race|Heard GAE* (seeing an African American Speaker and hearing GAE as compared to *Same Race|Heard GAE* (seeing a European American Speaker and hearing GAE). We also found a significant main effect of age, suggesting that older children looked more to the target images overall in this experiment. Lastly, all other main effects and interactions were not significant.

Table 4

Model estimates for generalized linear mixed effects model. The measure of vocabulary is the raw score from the NIH Toolbox PVT.

R syntax: cbind(LooksTarg, LooksNotTarg) ~ 1 + Age (centered) + Vocabulary (centered raw score) + Speaker Group (contrast coded) + Speaker Group * Vocabulary + Speaker Group * Age + (Speaker Group | participant) + (1| item)

Fixed Effects		***b***	**SE**	**z**	
Constant		5.11	.55	9.20 ***	
Age		1.38	.50	2.76**	
Vocabulary		-.40	.51	-.77	
Speaker Group (Contrast 1)		-.13	.83	-.15	
Speaker Group (Contrast 2)		-2.07	.70	-2.95 **	
Speaker Group (Contrast 1) *Vocabulary		-.99	.85	-1.17	
Speaker Group (Contrast 2) * Vocabulary		-.07	.69	-.11	
Speaker Group (Contrast 1) *Age		.13	.83	.16	
Speaker Group (Contrast 1) *Age		-.87	.69	-1.27	
Random Effects	**Parameter**	**Variance**	**SD**	**Corr**	
Participant	Intercept	5.5	2.3		
	Speaker Group (C1)	14.6	3.8	-.03	
	Speaker Group (C2)	10.1	3.1	-.03	.58
Item	Intercept	1.1	1.0		

*p <.001***, p < .01 **, p < .05* , p ≥ .05+*

Figure 3. Mean looks to target by age group (median split, months converted to years)

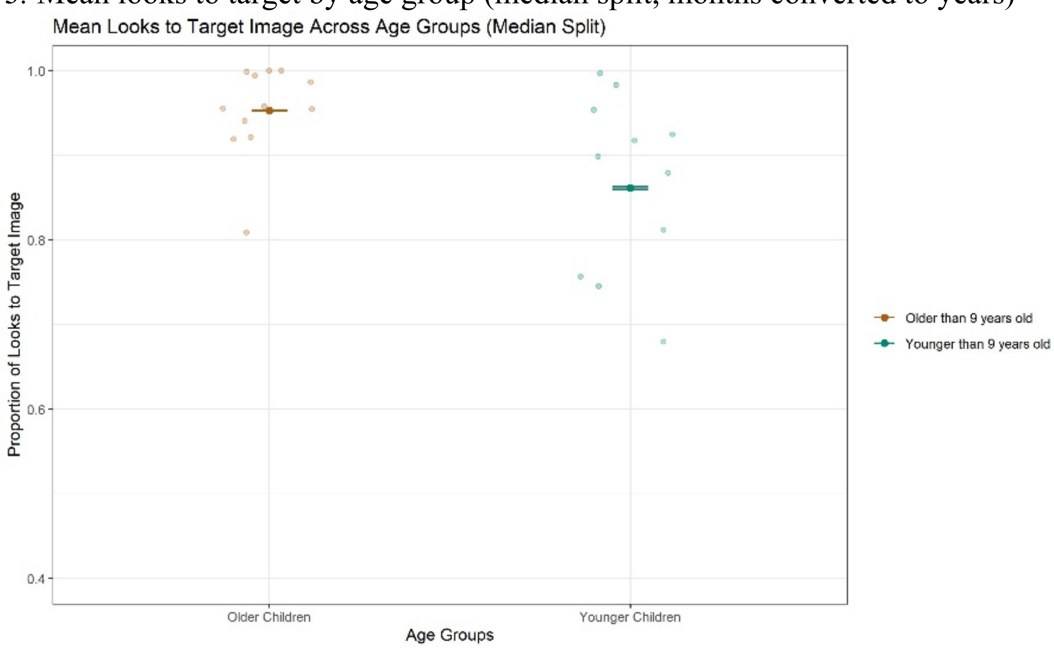
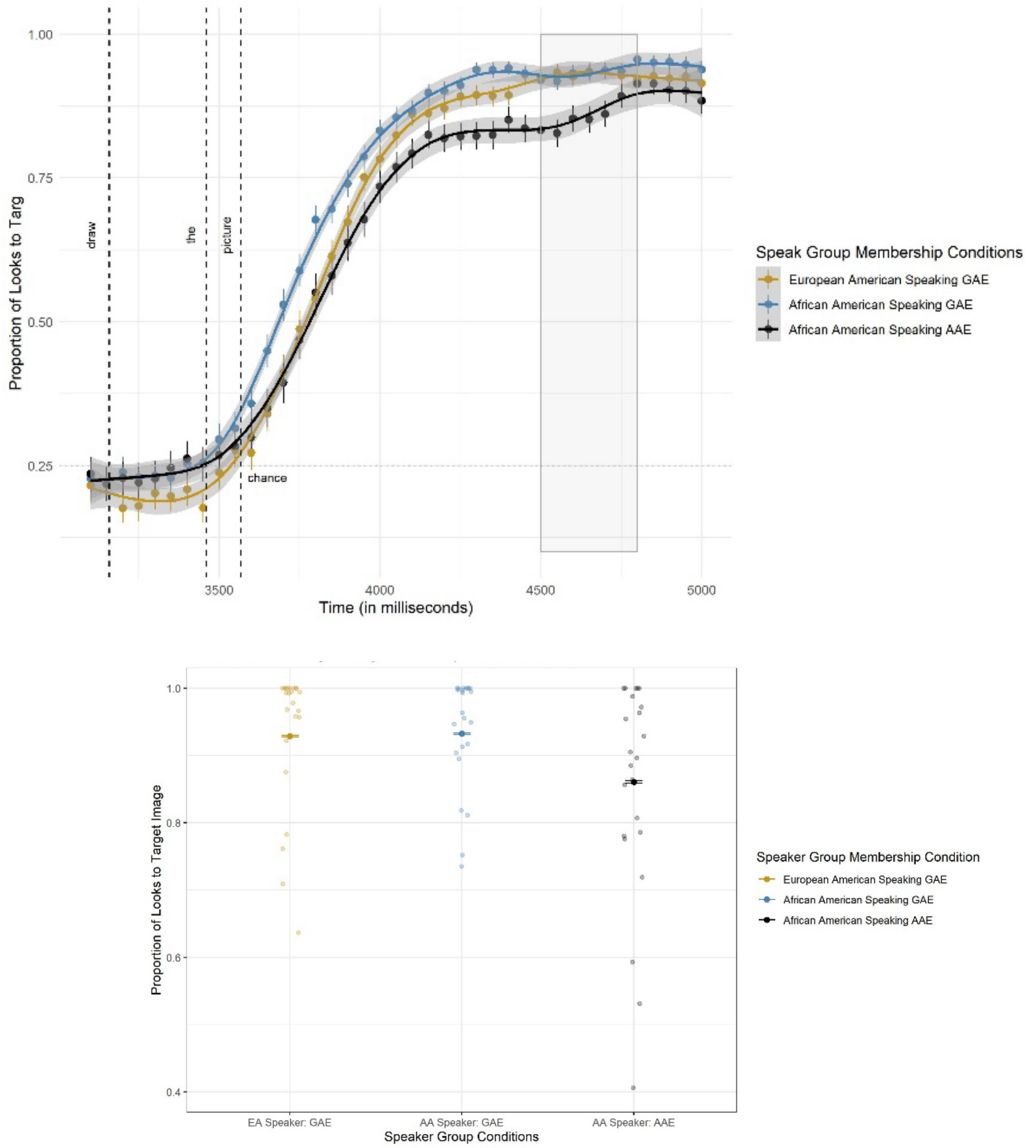


Figure 4. Looks to target over time as a function of speaker group condition (top plot) and looks to target in the analysis window (bottom plot).



Discussion

The results in Experiment 3.1 align with previous social psychology findings demonstrating that children of European descent attend to auditory linguistic cues over and above visual cues to ethnicity, even at an age in which children have shown emerging sensitivity to race and ethnicity (Aboud, 2003; Bar-Harim et al., 2006; Kelly et al., 2007, Shutts et al., 2009). Based on the Kinzler et al. (2009) study, one of our two alternative predictions was that children would exhibit a dialect familiarity benefit regardless of the speaker's racial background. First, we observed that GAE-speaking children showed efficient lexical processing in speaker group conditions in which listeners heard GAE paired with a European American face. Second, children showed more efficient lexical processing in speaker group conditions where the stimuli dialect matched the dialect of the participants (GAE), regardless of race. This differs from Weatherhead et al. (2018) which would have predicted a null result; listeners would demonstrate no difference across any of the three speaker conditions. Moreover, these effects were observed in sentential contexts where facilitating semantic information was present. This differed from those of Erskine (2022a) who found more efficient lexical processing in a familiar dialect for GAE-speaking children only in contexts without facilitating information (sentences with neutral verbs), a point we return to later in the discussion. Accompanying these results, we find evidence that children of European descent attend to signal-related cues over and above visual cues to ethnicity, even at an age in which children have shown emerging sensitivity to race and ethnicity (Aboud, 2003; Bar-Harim et al., 2006; Kelly et al., 2007, Shutts et al., 2009). The implications of these findings are discussed further below.

Among the European American children in this study, our results partially converged with previous studies on speech processing among infants (Weatherhead et al., 2018) for speakers of the same race. More specifically, children demonstrated efficient lexical processing for speakers of the same race and same dialect (i.e., hearing GAE from a European American Speaker). We did not have a condition to evaluate an unfamiliar dialect paired with the face of a European American, so it is hard to predict older children will fully replicate previous findings. However, we did find that listeners' sustained advantage in recognizing words in a familiar dialect (GAE) remained consistent despite differences in the speaker's ethnicity (i.e., hearing GAE from either a European or African American). These results are not surprising because we examined effects of speaker identity in a group of older children (6-11 years old) whereas Weatherhead et al. examined effects of race and accent in a group of infants (16 months old). Older children likely have more variable experiences with different groups of speakers than younger infants, whose interactions tend to be more restricted. Second, we recruited children from a relatively diverse environment (i.e., Prince George's County in Maryland) where listeners have opportunities to observe and interact with African American and European American individuals who use GAE in their everyday lives and across different social environments (e.g., in their local communities, at school, among adults and similar-age peers, etc). In this case, it is plausible that African American and European American faces served as familiar race conditions (see Weatherhead et al. 2018). Perhaps, socialized experiences with African American speakers assisted children in building stronger expectations about the speaker's language use. Further, the absence of differences between the two race conditions for GAE speakers is in line with other studies that have shown that European American listeners (5-7 years old) usually prioritize accent/dialect-related cues over and above visual cues about a speaker's race or ethnicity during sociolinguistic decisions about friendship (Kinzler et al. 2009) or speaker credibility (Kertesz et al. 2021). The studies on findings have been most reported for studies on social decision making in which listeners choose friendships (Kinzler et al. 2009) or determine the credibility of speaker, but we add to this body of literature by showing that these preferences extend to spoken language processing environments.

In contrast to Weatherhead et al. we did observe an effect of dialect variation among speakers of a different race. That is, children demonstrated less efficient comprehension of AAE than GAE, when the dialects were paired with the face of an African American. Similar to the arguments posed above, it could be the case that while European American listeners are familiar with African Americans using GAE, they are less familiar with African Americans who produce AAE. This explanation is plausible as sociolinguists (Green, 2002; Wolfram, 2001) have suggests that in environmental contexts where European Americans are present, African Americans who speak both AAE and GAE, will typically default to GAE forms. Moreover, we live in a society where dialect-shifting is implicitly encouraged and expected in many social contexts. Another reason that explains this finding is that we used a more demanding task than Weatherhead et al. Perhaps when the demands of spoken language processing are increased, children recruit speaker information in ways that change comprehension outcomes across studies.

The results of this experiment could be interpreted as providing support for either perceptual processing or socially enriched speaker contexts (i.e., the salience of speaker identity through linguistic and nonlinguistic speaker identity cues). The perceptual *processing* hypothesis focuses on the speech signal only and is agnostic about the role of socioindexical cues above and beyond the speaker's dialect. Children in the current study showed a listening preference for GAE, their primary dialect, regardless of the speaker's ethnicity. This is consistent with a perceptual processing hypothesis which predicts a processing advantage for the familiar dialect. Alternatively, this finding is also consistent with the predictions about the importance of the socially enriched contexts that provide both linguistic and non-linguistic information about the speaker's identity. A socially-enriched account posits that the listener's experiences with a variety of speaker-related cues and the salience of both nonlinguistic and linguistic cues serve as the driving force for ease of spoken language processing. In this experiment, we included two salient speaker identity cues, the dialect that was heard and the race/ethnicity of the speaker. The conditions that were most facilitating for children in this study included at least one variable (i.e., dialect) signaling the speaker's identity. Therefore, our results do not rule out the validity of perceptual or socially enriched frameworks that include the influence of nonlinguistic speaker identity cues trust frameworks for explaining relationships between dialect variation and the lexical processing efficiency.

One of the most intriguing findings from this study is that differences in lexical processing efficiency between the familiar and the unfamiliar dialect were observed in a semantically facilitating context (i.e., sentences containing predictive verbs). This contrasts with the results of Erskine (2022a), which found an advantage for the familiar dialect only in a neutral contrast. *What accounts for the difference in results across studies?* The crucial distinction between the present study and Erskine (2022a) is the inclusion of visual information about the speaker in the current study. It is plausible that while visual cues were not prioritized for children's social listening preferences, they might have had some effect in favor of amplifying the salience of speaker identity. Erskine (2022a) suggested that the reason effects of dialect variation were observed in neutral but not predictive contexts was because listeners had facilitating lexical information (predictive verb) that could be leveraged to enhance word recognition. This additional boost from the hearing words in the predictive context helped listeners overcome the challenge of dialect differences. Thus, dialect differences were more salient when higher-level contextual cues were reduced. The current study provided children with additional information (i.e., nonlinguistic information) that could have been used to highlight differences in listening contexts that were socially familiar or less familiar to the listener. Based on the results of Erskine (2022a), we should expect these findings to observe these findings also in neutral verb contexts and they might even be more pronounced because semantically neutral sentences provide fewer opportunities to integrate higher-level sources of linguistic information that support processing. However, due to the small number of items containing neutral verbs, we were unable to assess this prediction, but this could be considered in future studies using similar paradigms.

Another source of difference is the later window of analysis that was chosen by the cluster analysis in this study (1342 ms after verb onset). By contrast, Erskine et al. (2022a) used a window that began 585ms after verb onset. Somewhat slower processing speeds in general were observed in the current study: peak looks to the target for the *Same Race|Heard GAE* speaker group condition occurred at about 4500 ms after sentence onset, while peak looks to the target in Erskine et al occurred at about 4000 ms after sentence onset. This experiment included visual information and sociolinguistic effects are usually late arriving relative to linguistic factors which typically influence earlier stages of language processing. This aligns with theoretical notions of processing. Therefore, it is not surprising that speaker identity (or socioindexical) effects will occur much later during real-time linguistic processing because children are integrating visual and auditory information.

Additionally, the cluster analysis excluded the Speaker Group condition, *Different Race|GAE*. This is because the eye gaze patterns for this condition diverged from the gaze patterns for the other two conditions much earlier (at about 70 milliseconds prior to the onset of the noun, see Figure 2). A qualitative inspection of the audio recordings revealed that the speaker of the Different Race|GAE condition (i.e., hearing GAE from an African American Speaker) produced sentences with a very clear and steady voice and a highly resonant voice quality. Moreover, many children in the study reported that they liked listening to the African

American speaker producing GAE. Therefore, early arriving differences in this condition may reflect something unique about this speaker's voice such that many children preferred listening to this particular voice, independent of the experimental manipulation. For these reasons, data from this condition were excluded from the cluster-based permutation analysis. A limitation of this study which can be addressed in future research is that it included only one speaker for each of the three conditions.

In addition to effects of speaker identity, we replicated previous results demonstrating that spoken language processing improves as children age. It is unsurprising that we found a significant effect of age. The sample in Experiment 3 included a broader age range (6-11) than the previous study (5-8). This finding converges with early studies (Fernald et al. 1998; Fernald et al. 2006; Marchman & Fernald, 2008; among others) showing that children become more efficient language processors over time. This is most likely attributed to the growth and maturity of their linguistic and cognitive skills.

Remarkably, vocabulary was not a robust predictor of general processing or performance in any of the speaker group conditions. Moreover, there was an absence of statistical significance despite having adequate variability in children's scores (range=67-95, mean=83, sd=8). There are two possible reasons for this finding. First, this study used a measure of vocabulary that differed from prior experiments (Exp 1 and 2) and other studies (Borovsky et al., 2012; Mani & Huettig, 2012). Specifically, we used a screener provided by the NIH Toolbox as opposed to the PPVT-4 (Dunn & Dunn, 2007), an instrument that has been used in clinical contexts for its diagnostic validity in identifying children with lower levels of receptive vocabulary. We chose this measure to shorten the length of visits during the pandemic, which increased parents' overall comfort in participating in person. It is plausible that the NIH vocabulary subtest on its own is a less sensitive measure of vocabulary. Another concern could be related to the vocabulary used in the analysis (i.e., a raw score that evaluates the number of words known by a child as opposed instead of an age-normalized standard score). However, it should be noted that similar results were found when the standard score rather than the raw score was included in the analysis. Second, it is possible that effects of vocabulary were masked by the inclusion of age in the model. Statistically, age tends to have high predictive value because it is correlated with many aspects of children's cognitive and linguistic growth. Due to the wide age range that was sampled in the study, we included age to account for differences in children's performance that could have been due to overall development. However, this inclusion may have obscured contributions of vocabulary. In fact, this seems to be a likely explanation, because excluding age resulted in vocabulary becoming a significant predictor of overall linguistic processing. However, the best fitting model in this study included age and vocabulary, even though only age was a significant predictor.

In summary, Experiment 3.1 provides additional evidence that children regard socioindexical information in ways that are nontrivial for linguistic processing. Moreover, we found that when visual aspects of the speaker's identity are included in the stimuli, effects of dialect variation are not limited to neutral sentential contexts. Among European American listeners, who were primarily monodialectal speakers of GAE, we found consistent evidence that listeners prioritize dialect and not race as an index of the speaker's identity. For familiar speaker communities (i.e., African American or European American speakers), children established linguistic expectations for some speakers, which affected their sensitivity to dialect differences.

Experiment 3.2: African American Listeners

One question that arises from this finding is whether children across all sociocultural contexts with different socialized experiences will demonstrate similar spoken language processing preferences. More specifically, *what do we predict for listeners who are more likely to encounter speakers of dominant and minoritized dialects in their immediate environment?* The second experiment (Experiment 3.2) reported in this paper tackles this question directly by examining the effects of speaker identity on spoken language comprehension among African American school-age children. African American children in Prince George's Community are likely to have socialized experiences in which they hear AAE and GAE produced by other African Americans in their local communities. Additionally, children in this study are at an age in which they are also experienced with observing European American speaker produce GAE. Based on each listener's lived experiences, it is plausible that they have established expectations for speakers where there is racial overlap (i.e., African American faces) and where there are differences (i.e., European Americans). Unlike the predictions for experiment 3.1, the predictions for experiment 3.2 are not mutually exclusive.

1. We predict that **visual cues indicating the speaker's race will be prioritized over cues related to dialect use.** It is predicted that African American children will use race and ethnicity as a stronger cue to speaker identity than dialect because across most of their social experiences European American speakers use GAE as their primary dialect, whereas African American speakers variably use AAE and GAE. Therefore, race becomes a stronger identifier for reasoning about linguistic expectations.

2. We also predict that **listeners' dialect use will influence comprehension of words in contexts where either AAE or GAE is paired with an African American face.** The degree to which African American listeners process words efficiently in AAE or GAE for speakers of the same race (i.e., African Americans) will depend on the listener's experience and use of either AAE or GAE. Children who produce more AAE will show an advantage in processing words in AAE as compared to GAE. However, children whose primary dialect is GAE or who have limited experiences with AAE, will be less efficient at processing AAE paired with African American faces.

Listeners

Twenty-six African American children (16 female) between the ages of 84 and 132 months (mean=106, sd=14) were included in this study. Children were recruited during the height of the Omicron wave of COVID-19. To ensure an adequate sample size the geographic reach was widened to include children from states on the broader eastern coast of the United States. As a result, we included children from areas near New York, Washington, D.C., Maryland, and Virginia. An additional 10 children were recruited but were not included in the final sample for the following reasons: experimenter error or equipment failure (n = 6) and missing data that exceeded the threshold for inclusion (n = 4).

Children were recruited from three different places: a) the University-based consortium of families (n=11), b) a local Maryland chapter of the National Association for the Advancement of Colored People (NAACP-MD) (n=9) and c) a local church in New York (n = 6). The majority of participants included in this study were from families with higher levels of maternal education (e.g., received a college degree or more, n=22). To accommodate flexible participation in research, we tested children either virtually or in person using webcam eye-tracking (see Visual World Paradigm Procedures for further details). Although we included remote participants in our recruitment and testing, we ultimately included only a very small number of children who received this method (n=2). A total of six children initially participated in this study remotely, however four of these participants had a large amount of missing data due to attentional, equipment, or compliance difficulties. These challenges are inherent to remote eye-tracking research where there is less control over the testing environment and more opportunities for distractible environments.

The final sample included children with a history of typical speech or language development and a range of dialect density scores on the DELV-ST. Dialect density was calculated using a proportional score derived from the number of features produced in AAE relative to the number of AAE and GAE responses, a measure that has been used by several studies on dialect variation (Craig & Washington, 2004; NP Terry, Connor, Johnson, et al. 2016; NP Terry, Connor, Thomas-Tate, et al. 2010). Sixteen children were categorized as speakers of GAE, six children demonstrated *Some* Variation and four children demonstrated *Strong* Variation.

Table 5
Demographic summary of the 26 participants and their scores across standardized assessments administered in Experiment 3.2.

Female	16
Mean Age (in months)	106 (22) *range=72-138*
NIH PVT Toolbox *Mean Raw Score* (SD)	83 (8) *range=67-100*
Mean Standard Score (SD)	107 (15) *range=84-142*
Maternal Education	
Mid (some college)	1
High (college degree or more)	21
Declined to report	4
DELV-Screening Test (Part I)	
Speaker of GAE	16
Speaker with SOME or STRONG Variation	10

Mean Dialect Density Scores (SD)	.16 (.11) *range=.00-.44*

Visual World Paradigm

Auditory and visual stimuli. The same auditory and visual stimuli described in Experiment 3.1. were used.

Webcam eye-tracking. Data for the African American participants were mostly collected prior to data for the European American participants. As noted above, these data were collected during the height of the Delta/Omicron waves. Because families were reluctant to come into the lab for in-person visits, we decided to use online webcam eye-tracking. Ultimately, because of concerns about data quality, it was necessary to transition to in-person visits with African American participants, but we continued to use webcam eye-tracking so that we could combine data across the virtual and in-person visits. Previous studies have shown that webcam eye-tracking is a useful tool that can be used to evaluate linguistic phenomena among populations typically excluded from laboratory-centered research. Moreover, studies have shown webcam eye-tracking is able to replicate results that have been previously reported from studies in more controlled laboratory settings (Semmelmann & Weigelt, 2018; Venker et al. 2020; Yang & Krajbich, 2020). Nevertheless, webcam eye-tracking introduces additional sources of variability that can impact gaze data quality (i.e., challenges related to screen size and camera quality or increased opportunities for environmental distractions). To minimize these challenges, a number of precautions were taken. Participants were asked to sit in a quiet room with an adult nearby to support the remediation of technical challenges and the four images on the screen occupied coordinate positions that placed them close to each of the corners on the screen. These decisions were necessary to reduce the challenges introduced by these sources of variability and assist with later manual coding procedures.

The visual world paradigm was implemented using Penn Controller for Ibex (Zehr & Schwarz, 2018). A combination of customized java scripting and default functions provided by Ibex were used to create the trial sequence that was described in Experiment 3.1. In this section, we provide a brief description of the experimental sequence in Experiment 3.2. The experiment began with a welcome page that included an embedded consent form and instructions on how to set up the webcam. After completing informed consent, caregivers gave the browser permission to use the webcam which recorded children during the course of the experiment.

Once the camera was set up, children were seated in front of the computer screen at a relatively close distance to the child for comfortable viewing of the images. They were then provided with general instructions about how to complete the Galaxy Quest game. Like Experiment 3.1, children were informed that they would be introduced to specific commanders who would tell them something about one of the images on the screen. Children were instructed to listen carefully and look at the image that best matched the sentence that was heard.

Prior to meeting the "commanders" (i.e., the speakers assigned to each speaker group condition) and completing the 54 test trials, children completed eight practice trials to help familiarize them with the experiment and provide opportunities for manual eye-tracking coders to later determine whether gaze patterns were mirrored during the recordings. Additionally, because we needed to obtain accuracy during remote visits, verbal responses were collected instead of a touch response. To evaluate accuracy, pictures appeared on the screen with a border of a specific color that was fixed in each of the four quadrants. The borders were about an eighth of an inch, making them visually salient to participants. In other words, the top-left image always appeared with a red border, the bottom-left image had a green border, the top-right image had a yellow/orange border and the bottom-right had a blue border. During the first four trials, children saw two astronauts and two planets that were one of two colors, blue and orange. Children heard simple phrases "*Find the [color] planet*" and verbally responded with the color border around the image. This continued until children identified each of the four objects on the screen. The verbal response was used to capture accuracy on the task. The last four practice trials mirrored this procedure with new images, two spaceships and two astronauts, but instead of hearing object names embedded in simple carrier phrases (*find the*) children heard sentences that were comparable to the test trials *"He was watching his friend choose the [color] spaceship."* This was done to help children transition from the practice trials to the test trials which would be used for later analysis. All children demonstrated 100% accuracy on these initial eight practice trials.

After the practice trials, children were introduced to all three commanders one at a time, and heard a verbal introduction from each commander that was recorded in the dialect assigned to each speaker group condition (*"Hi. My name is Commander [name]. I'm going to tell you something about one of the pictures on the screen. Listen carefully!"*). The experimental trials then proceeded as described in Experiment 3. The critical difference between Experiment 3.1 and 3.2 is that during webcam eye-tracking, there was no fixation point to initiate the start of each new trial. Instead, after children provided a verbal response to identify the quadrant of the target image, they pressed the space bar to advance to the next trial.

In all other respects, the task followed the procedures used in Experiment 3.1. Children received the same four experimental lists that were described earlier with sentences across conditions intermixed into a single block of eye-tracking. However, instead of automatic eye-tracking, gaze data were recorded using a webcam.

Administration procedure. Children participated in this study either in person or remotely. For remote participation, caregivers were encouraged to have two electronic devices, a computer and an additional device such as a phone, computer, or tablet which could be used to open a Zoom link. Two devices were necessary so that children could complete the eye-tracking task on a computer and have a second device where the experimenter could be present to provide technical support or reinforcement about the child's posture or attention to the experiment. The day before a scheduled visit, parents received an email with instructions about how to join the experiment Zoom session and a link to the experiment. At the start of the virtual session, parents provided informed consent and used the instructions to set up the eye-tracking experiment with support of the examiner on the Zoom call. After the experiment was set up, children were seated in front of the computer close enough to the computer so that the webcam could record their eye movements in closer proximity. The entire experimental procedure took approximately 1 hour. Preparing the experiment took about 20 minutes, the visual world paradigm took 20 minutes, and the remaining time was dedicated to breaks and completing two additional language measures (i.e., vocabulary and language variation screeners). This procedure was used for a total of six children, but only two of the participants who participated remotely were included in the final sample. The exclusion of participants was due to equipment failure (i.e., some children had computers with malfunctioning speakers which created a noisy listening environment) or difficulties with compliance.

In-person visits took place at the university. To create environments that were similar to virtual visits, the same experimental procedures were used for the in-person visits. Informed consent was completed at the start of the session with the child's caregiver. Children were seated in a quiet room in front of a computer and positioned close to the camera. However, unlike virtual sessions, in-person visits required a single device because the experimenter was present to assist with setting up the task and providing additional instructions. While in-person visits were not the same as remote visits at home, these steps were taken to minimize the differences across administration procedures.

Data preparation and cleaning

Manual coding. All eye movements were coded from webcam recordings by trained research assistants using (Peyecoder, Version 1.1.5, Olson et al., 2020) at the rate of 30 frames per second. The coders were not able to hear the audio, view the display screen, and were not aware of the experimental condition manipulations, trial type or target images. These steps were taken to minimize the coders' bias to label looks as being towards the target referent. All images were coded as one of the quadrants (top left, top right, bottom left or bottom right), an area of the screen was not an AOI (i.e., center, offscreen, a blink, track loss). The *track loss* code was used any time the video was too pixelated or was otherwise disrupted making it difficult to reliably code children's gaze patterns. All images were coded relative to the coder's view of left and right.

Manual coding training and calibration. Before webcam videos were independently coded for analysis, research assistants underwent intensive training and calibration to become reliable coders. As described by Venker et al. (2020), the manual coding of eye-tracking experiments is by no means a trivial process. Hand coding requires extensive training, practice, calibration with a team of coders, and periodic review of the videos that have been coded.

At the onset of training, research assistants were provided with a detailed coding guide that described the coding process including information about the codes that would be used, the coding software and programs including the functions coders would use to code each video, and visual examples of *typical* looking behaviors that would be observed while coding. A lead coder (i.e., a person that was trained in using the coding software and had extensive experience with hand coding) was selected to provide research assistants with the scaffolding required to help them become reliable coders.

The training procedures contained three stages: a) guided coding, in which trainees coded trials in the presence of a lead coder following an answer key, b) independent coding with an answer key accessible, and c) independent coding without an accessible answer key. This process was used to gradually fade the level of support that was necessary for research assistants to code videos. It also allowed coders to smoothly transition from training to the calibration phase, which we discuss in greater detail below.

During guided coding, trainees met in small groups with a lead coder. As a collective the lead coder and trainees coded six example trials with an answer key visually available. This was done to ensure that team members learned about the application of specific codes and how different looking patterns were mapped to specific codes. During these meetings, the lead coder provided additional guidance and feedback as often as necessary. In the second phase of training, trainees were provided with a new set of example trials which they coded independently (i.e., without a lead coder) with an answer sheet. This procedure provided opportunities for coders to engage in independent learning and practice of skills acquired in the first few meetings with the lead coder. Questions related to trials coded during this phase were addressed in a separate group meeting with the lead coder. In the last training phase, coders were provided another set of trials which they coded independently, but this time without access to an answer key. This step was crucial for providing trainees with opportunities to practice coding with less support. During this stage, trainees coded videos and then turned in their responses to the lead coder who would evaluate their reliability against an answer key and provide additional feedback and guidance as deemed necessary. The reliability of coders was also examined for inter-rater reliability against an answer sheet. Inter-rater reliability was calculated using a default function provided by Peyecoder. In depth details about the interrater reliability calculator can be found in the coder manual accompanying the software (link here) (Olson et al., 2020).

To calculate interrater reliability among coders, we used the 'Compare against" function in Peyecoder which calculates three specific measures: frame agreement, comparable trials, and shift agreement. Frame agreement corresponds to the general agreement between coders about gaze location (i.e., the percentage of frames in which two or more people coded the same response). This index provides a measure of consistency or general agreement about where children are looking on the screen. Comparable trials indexes whether manual coders have coded the same number of responses. Lastly, shift agreement evaluates the percentage of agreement across time stamps of events coded with the same response within close proximity of each other. That is, it examines the extent to which coders are consistent in how the number of frames that a response and whether the timing of this response is in close proximity. A coder was deemed reliable if they demonstrated at least 80% agreement for at least 2 of 3 of the measures. This criterion was used because the "Compare against" function was originally designed to be used with looking while listening paradigms, in which gaze patterns are monitored in only 2 to 3 areas of interests (usually, left, right, and center for fixation points) (Venker et al., 2020). In the eye-tracking task in this study, we monitored four relevant areas of interests (top-left, top-right, bottom-left and bottom-right) on screens with variable sizes and dimensions. Usually, in task with more than two area of interests, looks can be more difficult to assess, especially gaze movements in the up and down position. As a result, we used a cuttoff level for interrater reliability that is lower than previous studies (i.e., 80% agreement).

Once trainees reached the criterion level of 80% interrater reliability in at least 2 measures (against the answer sheet), they were advanced to the calibration phase. If trainees did not reach this criterion level, they were assigned additional trials until the necessary agreement threshold was reached. The training process with four coders took approximately 480 hours to complete.

Calibration and post calibration review. During the calibration process, trainees were assigned 20 calibration files across two participants from two different eye-tracking experiments. Trial assignments were based on a set of files that had been previously coded by trained coders. Additionally, answer keys were developed for each of these files and were used to evaluate the calibration files completed by the trainees. For calibration files, trainees independently coded all assigned trials. These trials were later assessed for reliability by the lead coder using the "Compare against" function. If coders reached the predetermined 80% threshold in two of three measures, they were assigned critical trials from the experiment reported in this paper. If coders did not reach the predetermined criterion, they were assigned a new set of previously coded trials. Trainees remained within the calibration phase until they met the criterion threshold. All coders met the criteria for inter-rater reliability within two attempts. After coders completed training and calibration and coders advanced to independently coding videos, the lead coder periodically reviewed files to examine the quality of independently coded data. To ensure the quality of data, the lead coder would randomly evaluate an early and middle portion of trials from videos that were in progress of being coded. If glaring issues were observed, the assigned coder was asked to recode the video. If no concerns were noted by the lead coder, the coder assigned to the video continued to code the video until it was completed. This was usually done early on to avoid recoding of entire participant. *Data preparation and cleaning.* Peyecoder automatically extract all codes and converts information from frames to 33ms bins of time. Prior to extracting all of the data across participants, all test trials were independently examined by an independent coder and compared to “practice trials” to determine if recordings were mirrored. Using a built-in Peyecoder function, codes were transformed to accommodate mirrored recordings.

Data cleaning and screening followed the same procedure described in Experiment 3.1. Children were excluded from all further analyses if they had fewer than 50% of reliable trials in a single block of eye-tracking. The quality of the data was then examined in the time region of 3100 to 6000ms, the region beginning at the onset of the critical verb (*read/choose*) to 2000ms after the onset of the target referent (*book*). We did not make any adjustments to the window of time because the auditory stimuli used in Exp 3.1 were also used in this experiment, which included 1000ms of silence padded to both the beginning and end of audio files.

Data cleaning procedures resulted in the retention of 792 out of 854 trials. Six additional children were tested, but they had greater than 50% of missing data in at least one block of eye-tracking and were therefore excluded from all subsequent analyses. Twenty-six children were included in all analyses of eye gaze patterns. The proportion of missing data across participants did not differ across the experimental conditions, $F(2, 1047) = .296$, $p = .74$. The tables below provide a brief summary of the track loss analysis for participants belonging to each linguistic group before and after data cleaning.

Table 6

Summary of the track loss and sample analysis before trial-exclusion. AA is the abbreviation for African American and EA is the abbreviation for European American.

Speaker Group	Number of Trials	Mean Samples	Percent Track Loss (%)
Saw AA Speaker: Heard GAE	286	161	14.8
Saw AA Speaker: Heard AAE	284	161	14.6
Saw EA Speaker: Heard GAE	284	162	15.0

Table 7

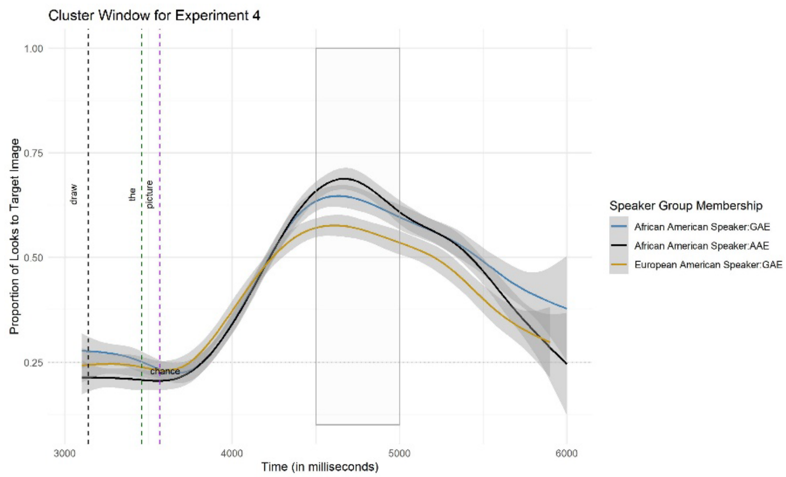
Summary of the track loss and sample analysis after trial-exclusion. AA is the abbreviation for African American and EA is the abbreviation for European American.

Speaker Group	Number of Trials	Mean Samples	Percent Track Loss (%)
Saw AA Speaker: Heard GAE	264	160	9.4
Saw AA Speaker: Heard AAE	263	161	9.3
Saw EA Speaker: Heard GAE	265	161	10.3

Results

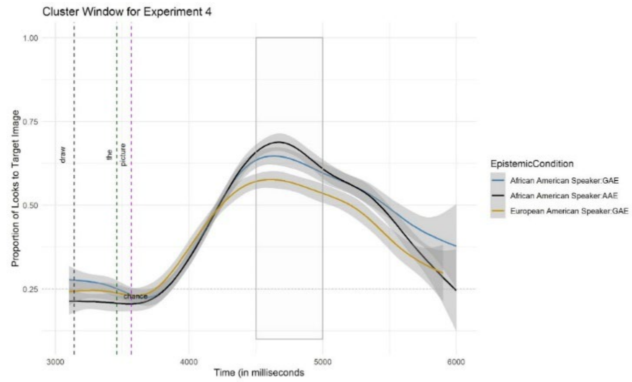
Cluster analysis procedures and results. The data were first restricted to a general window of time that began at 3100ms and ended at 6000ms. This window reflects the onset of the critical verb to 2400ms after the onset of the target referent. Like Experiment 3.1, this window was informed by empirical evaluation of the data in which looks hovered at chance across all three conditions until the period after the noun started to unfold (see Figure 5) and evaluation of the previous study in which effects were observed in a later window than found in prior studies (Blomquist et al., 2021, Erskine, 2022a).

Figure 5. Looks to the target image at an onset region that is 700ms prior to the onset of the critical verb (at 3100 ms) and an offset region at 6000ms.



The same procedures described in Experiment 3.1 were used to fit a mixed effects model predicting the empirical logit of fixating to the target image as a function of the fixed effect, Speaker Group and by-subject random intercept was fit to the data. The cluster-based permutation identified ten contiguous 50ms time bins from 4500ms to 5000ms in which the main effect of Speaker Group met the criterion threshold for significance. This is similar to the window of analysis in experiment 3.1 (4500 to 4800 ms). Figure 6 depicts the window of time identified. To be consistent with the analysis in experiment 3.1, we excluded the *Same Race|Heard GAE* speaker group condition in the cluster analysis, but we included it for visualization purposes in the figure below.

Figure 6. Looks to the target image in the region of analysis identified by the cluster analysis (Experiment 3.2).



Generalized linear model analysis and results. To examine the effect of socioindexical speaker identity cues on spoken language comprehension, we fit a logistic mixed effects model to the region identified by the cluster analysis (i.e., 4500 to 5000ms). The model included independent variables of Vocabulary (PVT uncorrected raw score), Age (in months), Dialect Density (proportion NMAE features on DELV-ST), Speaker Group (all three conditions), and pairwise interactions between Speaker Group and the aforementioned independent predictors. Also, while African American children with a range of dialect density scores were recruited for this experiment, the number of participants was too small to separate into linguistic groups (GAE vs. AAE speaker) for further analysis. Rather than including linguistic group, the model included Dialect Density to capture processing differences across the three speaker group conditions as a function of dialect use. All continuous independent variables were centered and the categorical predictor Speaker Group was contrast coded using simple coding. Here, the intercept corresponds to the grand mean of the empirical logit of fixating on the target image for children of average age and vocabulary. In the fitted model, the first contrast (Speaker Group Contrast 1) is interpreted as the difference in looking at the target image when seeing an African American Speaker and hearing GAE (*Same Race|Heard GAE*) from a reference level of seeing an African

American Speaker and hearing AAE (*Same Race|Heard AAE*). The second contrast (Speaker Group Contrast 2) evaluates the difference in looking at the target image when seeing a European American Speaker and hearing GAE (*Different Race|Heard GAE*) from the reference level. This allowed us to directly evaluate the hypothesis that children from African American backgrounds would be more likely to prioritize race-related cues above cues corresponding to the dialect spoken by the speaker. Dialect density was included in the model because, unlike experiment 3.1, there were a relatively large range of dialect density values. Lastly, we did not include maternal education as a predictor because children mostly came from families with higher levels of maternal education, only two families reported having maternal education levels that differed from a college degree (i.e., completed some college).

We found an interaction between Dialect Density and the second contrast of Speaker Group ($\boldsymbol{\beta}$= -.51, se= .25, z=-2.02, p <.05). That is, as dialect density increased, there were fewer looks to the target image in the condition *Different Race|Heard GAE* relative to the condition *Same Race|Heard AAE*, as illustrated in Figure 1. We also found a significant interaction between Age and the second contrast of Speaker Group ($\boldsymbol{\beta}$= .58, se= .29, z=-2.00, p <.05) suggesting that as age increased, children showed smaller differences in target looks between the following conditions: *Different Race|Heard GAE* and *Same Race|Heard AAE*, as illustrated in Figure 2. Finally, we found a significant interaction between Vocabulary and the second contrast of Speaker Group ($\boldsymbol{\beta}$= -.57, se= .29, z= -1.96, p <.05). As illustrated in Figure 3, an increase in vocabulary was associated with fewer looks to the target image for the Speaker Group condition, *Different Race|Heard GAE.* Lastly, we observed a marginal trend for the main effect of the second contrast for Speaker Group suggesting that children demonstrate fewer looks to target images in the Speaker Group condition, *Different Race|Heard GAE* relative to the condition

Same Race|Heard AAE. All other main effects and interactions were not significant.

Table 8
Model estimates for generalized linear mixed effects model. The measure of vocabulary is the raw score from the NIH Toolbox PVT.

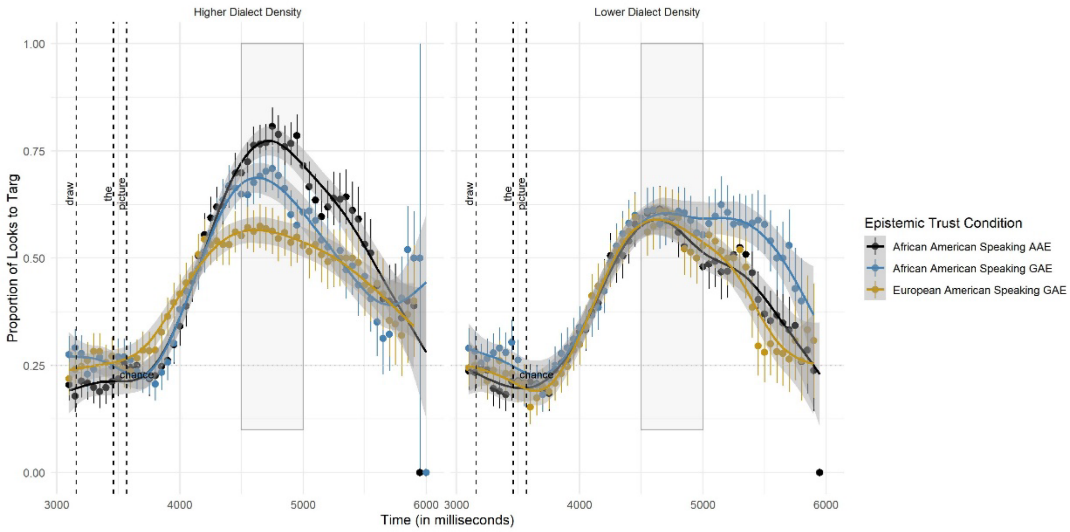
R syntax: cbind(LooksTarg, LooksNotTarg) ~ Dialect Density (centered) + Vocabulary (centered raw PVT score) + Age (centered) + Speaker Group (contrast coded) + Speaker Group * Dialect Density + Speaker Group * Age + Speaker Group * Vocabulary + (Speaker Group | participant) + (Speaker Group | items)

Fixed Effects	***b***	**SE**	**z**
Constant	.59	.22	2.70 **
Dialect Density	.29	.18	1.65+
Vocabulary	.30	.20	1.50
Age	-.32	.20	-.1.61
Speaker Group (Contrast 1)	-.34	.38	-.89
Speaker Group (Contrast 2)	-.76	.45	-1.69 +
Speaker Group (Contrast 1)*Dialect Density	-.40	.33	-1.20
Speaker Group (Contrast 2)*Dialect Density	-.51	.25	-2.01*
Speaker Group (Contrast 1)*Age	.23	.37	.62
Speaker Group (Contrast 2)*Age	.58	.29	2.00*
Speaker Group (Contrast 1)*Vocabulary	-.57	.38	-1.50
Speaker Group (Contrast 2)*Vocabulary	-.57	.29	-1.96*

Random Effects	**Parameter**	**Variance**	**SD**	**Corr**	
Participant	Intercept	.73	.85		
	Speaker Group (Contrast 1)	2.47	1.57	.39	
	Speaker Group (Contrast 2)	1.41	1.18	.10	.38
Item	Intercept	.36	.59		
	Speaker Group (Contrast 1)	.84	.92	.12	
	Speaker Group (Contrast 2)	2.47	1.57	-.29	.73

*p <.001***, p < .01 **, p < .05* , p ≥ .05+*

Figure 7. Interaction between Dialect Density and the contrasts of Speaker Group across time (top plot) and averaged in the cluster window as a median split (bottom plot).



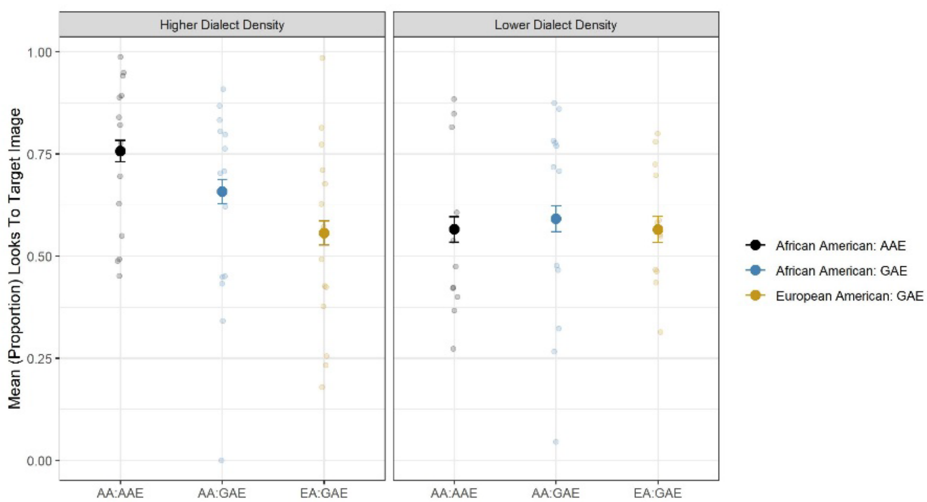
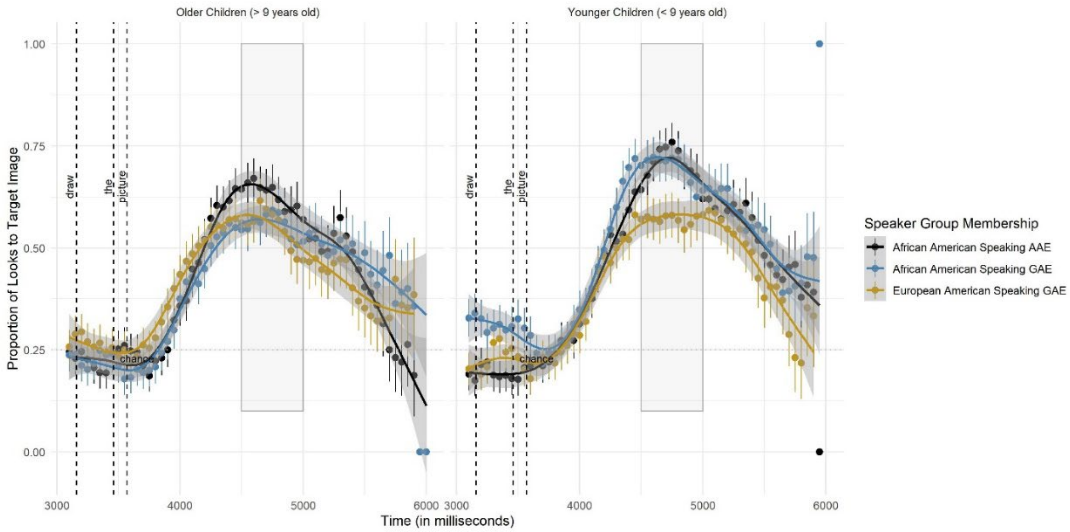


Figure 8. Illustrates the interaction between age and the Speaker Group 2. Age is visualized as a median split (older vs. younger children) but treated as continuous in the regression model.



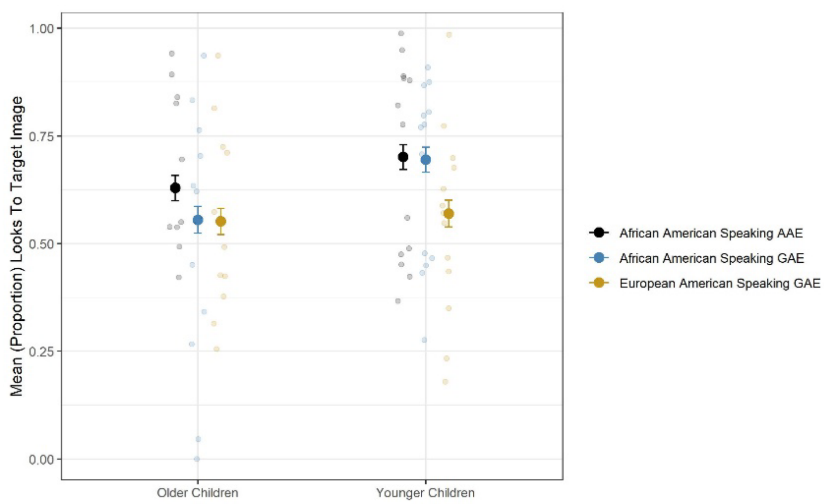
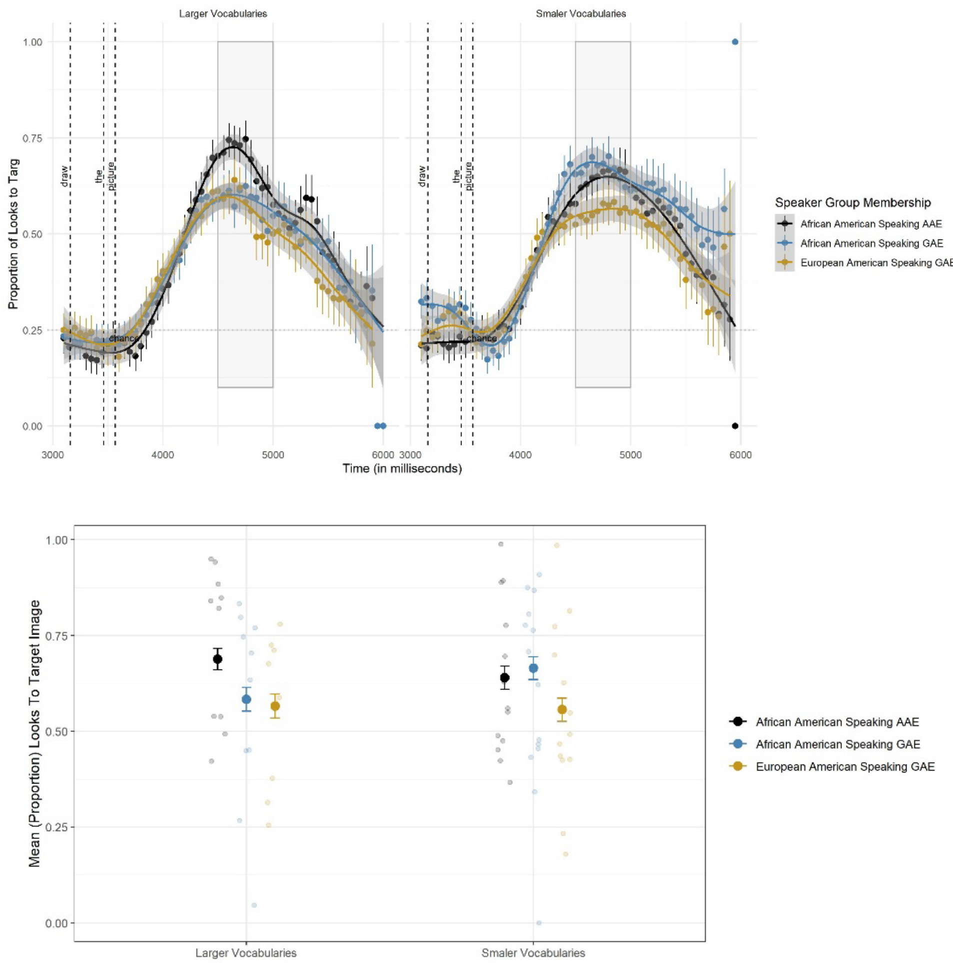


Figure 9. Illustrates the interaction between vocabulary and the contrasts of Speaker Group. Vocabulary is visualized as a median split (larger vs. smaller vocabularies) but treated as continuous in the regression model.



Discussion

Although a trend in the data emerged in which children showed more looks to the target image for African American speakers regardless of dialect, the overall pattern of results for experiment 3.2 is that the effect of speaker group on spoken word recognition for African American children depends on their overall use of AAE. Children with higher dialect density showed the most efficient lexical processing when they saw an African American face and heard AAE. Additionally, these same children showed the least efficient lexical processing when they saw a European American face and heard GAE. Children with lower dialect density, by contrast, showed no substantial differences in lexical processing efficiency across any of the three speaker group conditions.

Among higher dialect density listeners, we observed an overall advantage in comprehending words from individuals of similar racial background and dialect, but not on the basis of dialect alone (i.e., speaking GAE regardless of speaker race). This is supported by the finding that children with higher dialect density showed an increase of looks to target images when observing African Americans speak AAE and fewer looks overall to European Americans speaking GAE. Listeners with low levels of dialect density (i.e., African American children who speak GAE) showed similar lexical processing efficiency across all three speaker group conditions. Lexical processing efficiency was not based on either the dialect or race of the speaker. Based on informal conversations with parents, it is evident that children with lower dialect density potentially comprise two groups. Some children are primarily exposed to GAE with limited exposure to AAE from grandparents or other family members/friends, while other children are bidialectal speakers who speak AAE at home and GAE in formal settings such as the laboratory setting where data were collected. Since the DELV-ST measures production only in a formal setting, it is limited in its use to evaluate whether children are bidialectal. Interestingly, the results did not support the prediction that children who produced fewer AAE features would also be less efficient at processing AAE. Given that both groups of GAE-speaking African American children have some exposure to AAE, it is not surprising that they performed similarly in the two conditions with African American faces and different dialects (AAE and GAE). Similarly, since African American children were recruited from local communities with diverse speaker communities, they are likely to interact with a variety of speakers of different races or ethnicities who speak GAE.

Lastly, this study observed a trend in which race was a stronger overall cue than dialect. That is, this study found a marginal effect of the second contrast of Speaker Group. That is, there was a trend in which children were more efficient overall in processing words when AAE was paired with an African American face relative to when GAE was paired with a European American face. Additionally, the first contrast of Speaker Group was not statistically significant suggesting that children were equally proficient at processing words in AAE or GAE when paired with an African American face. Moreover, qualitative inspection of the data suggests that this trend may have been driven by children with higher dialect density. The absence of an effect may be due to a lack of power. While our sample size was comparable to previous studies on real-time comprehension of variation (Creel, 2012, Law et al., 2017)

, we used webcam eye-tracking instead of the automatic eyetracking used in these previous studies. A growing body of research has shown that while previously established effects can be replicated with webcam eyetracking, they are usually smaller in size than the average effects reported from in-person laboratory studies using automatic eye-tracking. This finding is usually attributed to increased variance (i.e., noise) within and across participants. Notably, studies that are able to detect smaller effect sizes have typically involved much larger samples (e.g., at least 50 participants) or more trials (Anwyl-Irvine et al., 2020; Venker et al., 2020; Yang & Krajbich, 2020). Perhaps with a greater number of trials and participants, the observed trend in which children with higher dialect density appear to prioritize race-related cues above accent would be significant. This limitation of webcam eyetracking also calls into question whether it is an appropriate solution for addressing barriers to representation and inclusion in research where effects are relatively small in size, even though it assists with the recruitment of children from under-represented backgrounds.

In summary, GAE-speaking African American children raised in diverse communities, like the children in this study (Prince George's County, MD), are likely to interact with both African American and European American speakers of GAE and with African American speakers of AAE. This could explain the performance of GAE-speaking African American children in this task where they show very little difference in lexical processing efficiency across speaker group conditions. In contrast to the results of Weatherhead et al. (2018) we did not observe evidence that for speakers of the same race, children will show poorer comprehension of words in the less familiar dialect. Children with high or low dialect density did not demonstrate significant differences in their comprehension of words in facilitating sentence contexts when listeners see an African American face and hear either AAE or GAE. As noted above, it is possible that all dialect/face conditions were familiar to the GAE-speaking African American children in this study. Thus, these findings for African American listeners are interesting because they suggest that their lexical processing efficiency is primarily driven by their experiences with different identities of speakers including information about the speaker's racial background and the dialect that was spoken.

Lastly, this study observed significant interactions between Speaker Group and Age and Speaker Group and Vocabulary. As age increased, the effect of Speaker Group decreased. This result may reflect children's lived experiences in which younger children, relative to older children, have fewer opportunities to interact with different kinds of speakers. Additionally, it may also reflect their maturing linguistic flexibility; as children develop more adult-like linguistic systems, they are better able to process speech from an unfamiliar speaker. The interaction between vocabulary and condition is unexpected and not entirely clear. As vocabulary size increased, the effect of speaker condition increased. It is not clear why acquiring more words would create larger differences in how listeners process AAE speakers who are African American and GAE speakers who are European American. If vocabulary was indexing something about GAE-alignment, then we would expect the opposite pattern. Additional analyses that included standard vocabulary size instead of raw score and included or excluded age yielded a similar result.

General Discussion

To summarize, the objective of this experiment was to investigate children's sensitivity to socially informed speaker-listener contexts. That is, we assessed spoken language comprehension of words in semantically facilitating sentences across a set of stimuli that varied as a function of linguistic (dialect) and non-linguistic (race) speaker properties. Among two linguistic communities, monodialectal European GAE-speaking children and African American speakers who varied in their use of AAE and GAE, we observed that speaker identity plays an important role in how listeners process language, but the cues that are prioritized by each listener group potentially differ. These differences may be driven by linguistic experience as well as more general familiarity with different kinds of speakers. Overall, this study provides additional evidence that, among 6-to-11-year old children, speaker identity cues do influence spoken language comprehension, especially when listeners are less familiar with the speaker's dialect or speakers that use those particular dialects.

First, this study provides evidence that language variation (i.e., dialect or accent) is a robust cue to speaker identity among children of European American descent. In Experiment 3.1, children demonstrated efficient word recognition for individuals producing a familiar dialect relative to a less familiar dialect, regardless of the speaker's race. This is consistent with previous research (Kinzler et al., 2009) that GAE-speaking children prioritize dialect over race in social contexts. The results were more complex for Experiment 3.2, in part because African American children showed more variation in nonmainstream dialect use than did the European American children. African American children who produced high rates of AAE (i.e., AAE speakers) generally showed more efficient comprehension of speakers of the same race and dialect. By contrast, African American children who produced low rates of AAE (i.e., GAE-speaking African Americans) did not show comprehension differences as a function of either race or dialect. As discussed above, the GAE-speaking African American children in Experiment 3.2 likely had experience with African American speakers of both GAE and AAE, and with European American speakers of GAE. These results indicate that lived experience plays a prominent role in linguistic processing.

We conducted two experiments that tested the validity of the perceptual analysis hypothesis in explaining relationships between dialect variation and spoken language comprehension. The results of Experiments 3.1 and 3.2 do not provide a clear answer. In Experiment 3.1, European American GAE-speaking children showed more efficient lexical processing for speakers of their primary dialect, regardless of race. This result is consistent with the perceptual analysis hypothesis and also with frameworks that emphasize the importance of socioindexical cues that extend beyond the auditory signal. This latter is true when dialect salience was amplified by including visual cues to the speaker's identity. Given the previous findings of Erskine (2022a), we found stronger effects of dialect variation in contexts that were previously impermeable to dialect variation (predictive verb contexts). However, these differences emerged only when visual identity of the speaker was present. In Experiment 3.2, AAE-speaking African American children were generally more efficient in processing words in speaker group conditions where the speaker was African American and spoke AAE. This result is consistent with both the influence of speaker identity and perceptual processing frameworks. If AAE-speaking African American children had also showed more efficient lexical processing for the African American speaker of GAE, relative to the European American speaker of GAE, this would indicate a clearer preference for race, regardless of dialect. Such a result would support the influence of nonlinguistic speaker identity cues over the perceptual processing hypothesis. However, this interaction was not significant, although there was a trend in the predicted direction. As noted above, Experiment 3.2 had lower power than Experiment 3.1 because of its use of the noisier technique, webcam eye-tracking instead of automatic eye-tracking.

In general, these results indicate that children demonstrate greater certainty in recognizing words produced by speakers who they are exposed to or interact with in their environments. For the European American listeners in this experiment, this includes speakers of the most familiar dialect as they are typically in environments in which speakers of diverse racial backgrounds generally use the dialect that is standard in the United States (i.e., GAE). For African American listeners with higher dialect density, we observed a strong preference for African American speakers of AAE followed closely by African American speakers of GAE. The least optimal listening context appeared to be European American speakers of GAE. This reflects environments in which they likely interact most frequently with African American speakers of AAE and least frequently with European American speakers of GAE except for formal learning contexts like school. Similarly, African American listeners with lower dialect density most likely interact with speakers of GAE more frequently across different racial backgrounds and are also likely to hear AAE spoken by other African Americans (e.g., grandparents, other family members, friends).

While the results of these two experiments provide evidence that nonlinguistic socioindexical cues that highlight differences across speakers influence spoken language processing of different dialects, they did not provide strong support for socially enriched speaker contexts over and above the perceptual analysis hypothesis. Despite the inability to differentiate these two frameworks, the need to find hypotheses that can capture differences across speakers that extend beyond the auditory/linguistic signal remain.

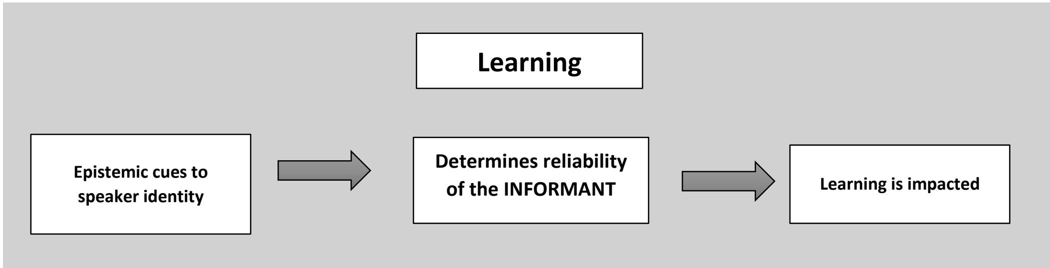
Reasoning about the speaker is guided by social nuances related to dialect variation. Several studies have shown that listeners use direct observations about a speaker's identity to inform inferences about whether the speaker and listener use language and convey meaning in similar ways (Weatherhead et al., 2021; Weatherhead & White, 2018; Shafto et al., 2012; Hay et al., 2006). Moreover, the results of this study supports previous research in showing that listeners' integration of speaker identity results in complex outcomes; spoken language comprehension is either enhanced or further diminished. As sociolinguists continue to advance dialect variation as a contributing factor for academic achievement, is important to understand *how* dialect does and does not influence aspects of achievement and *why*.

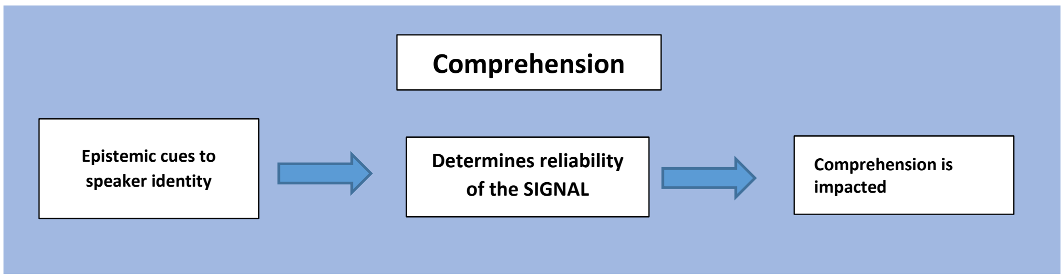
A lot of studies, including this one, provide useful evidence to tackle the *how* question, but rarely address the *why*. Rapid and efficient comprehension is influenced both by the clarity or familiarity of the auditory signal and the listener's expectations about what will be said. We reviewed earlier evidence that when assumptions about the speaker are violated, comprehension is negatively influenced (Babel & Mellesmoen, 2019; Babel & Russel, 2015; McGowan, 2015). The social contexts that inform the assumptions generated by the listener also rely on visual aspects of the speaker (race/ethnicity, gender, age). The perceptual analysis hypothesis may be indirectly capturing some of the non-linguistic aspects of the speaker because socioindexical information is also entrenched in the auditory signal. However, this study among several others highlight a need for accounts that will capture these elements more directly.

In the learning and computer sciences literature, some researchers have advanced ***the epistemic trust hypothesis*** as a way to understand what features of the speaker's identity (i.e., epistemic cues) serve as informative cues for learning (Shafto et al., 2012). Epistemic trust provides a theoretical framework to explain the task children face in learning from different individuals about the world around them. From the perspective of the learner, it is necessary to differentiate individuals who provide reliable or trustworthy information from individuals who are less reliable or provide less relevant information (Mascaro & Sperber, 2009; Pasquini et al., 2007; Corriveau, Meints & Harris, 2009b; Corriveau, Fusaro & Harris, 2009a; Corriveau & Harris, 2009a, 2009b; Clement et al., 2004; Harris & Corriveau, 2011; Jaswal et al., 2010; Jaswal & Neely, 2006). To filter out potential informants, the epistemic trust hypothesis posits that learners must identify relevant epistemic cues that determine whether a person is likely to be a reliable informant or not.

Studies from this literature have shown that children tend to show a preference for "familiarity" during learning. That is, previous studies have shown that children prefer to learn new words from caregivers as opposed to strangers (Barry-Anwar et al., 2017) and they will choose to be friends with or learn novel toy functions from speakers with familiar or native accents as compared to less familiar/non-native accents (Kinzler et al., 2009, Kinzler et al., 2011). These results are consistent with reports of a *familiarity* preference among children in the spoken language comprehension literature. For example, prior studies have shown that words are easier to recognize when they are produced by caregivers as compared to strangers or by individuals with familiar/native accents as compared to non-native accents. While the epistemic trust framework has not yet been applied to the literature on spoken language comprehension, it is plausible to predict that epistemic cues to speaker identity can be used to guide comprehension, just as prior experiences with different speakers (and the cues that inform assumptions about speakers) guide learning and other social decisions. Whereas epistemic cues help determine the reliability of the speaker for learning, we propose that these cues may be used by the listener to discern the reliability or familiarity of the signal which in turn guides the ease or difficulty in spoken language comprehension. This is illustrated by the figure below which compares how cues might be anticipated to influence comprehension as opposed to word learning. For learning (Figure 10, top figure), the first box represents speaker identity cues that are extracted in real time by the learner (e.g., language, accent/dialect, age, gender, race/ethnicity, etc). This information then feeds forward to facilitate internal reasoning processes about the speaker to determine whether the individual is a reliable informant or provides credible knowledge. Lastly, the decision about the informant's status will then directly influence whether children learn from a person or how challenging it is to learn from that person. We posit that the processes guiding learning will also be observed in a spoken language comprehension context. The critical difference might be that instead of considering the speaker credibility, listeners instead reason more directly about the signal itself (i.e., whether the individual will communicate information in ways that easily accessible or comprehendible). Future research could evaluate the claims made in the bottom half of Figure 10 by designing studies that directly pit predictions of the perceptual processing hypothesis against the epistemic trust hypothesis.

Figure 10. A comparison of the effects of epistemic trust on learning (top figure) and spoken language comprehension (bottom figure).


Learning
Epistemic cues to speaker identity
Determines reliability of the INFORMANT
Learning is impacted


Comprehension
Epistemic cues to speaker identity
Determines reliability of the SIGNAL
Comprehension is impacted

Chapter 4: Isolating effects of sociolinguistic cues on communicative intent

Over thirty decades of research at the intersection of education and sociolinguistics have provided consistent evidence that there is a negative correlation between nonmainstream dialect use and academic achievement. These findings have been interpreted to suggest that the long-standing gap in academic performance between African American students and their European American counterparts can be partly attributed to linguistic differences between African American English (AAE) and General American English (GAE). In support of this claim, a large body of evidence has found that African American children who produce high rates of AAE (and demonstrate little evidence of dialect-shifting) have lower scores on norm-referenced and standardized assessments of reading, mathematical reasoning, and spelling (Charity et al., 2004; Gatlin & Wanzek, 2015; NP Terry, Connor, Johnson, et al., 2016; NP Terry, Connor, Petscher, et al., 2012; Washington et al., 2018). More recently, some studies have gone beyond the assessment of decoding, spelling and reading comprehension (i.e., read and writing) and have focused on how dialect variation impacts processes that more general to children's learning environment, such as spoken language comprehension.

To date, investigations on the relationship between spoken language comprehension and dialect have shown that linguistic differences between AAE and GAE disrupt spoken language comprehension for AAE-speaking students. Edwards et al. (2014) examined the comprehension of words that were lexically ambiguous or unambiguous among four- to eight-year-old AAE-speaking children. For example, words like *gold* would be considered ambiguous as they could be perceived as either *goal* [goʊl] or *gold* [goʊld] among AAE speakers. Edwards and colleagues found that AAE-speaking children were overall less accurate in comprehending words that were lexically ambiguous relative to unambiguous words that did not differ between AAE and GAE (e.g., *bus*). Earlier studies on the comprehension of third person singular -s show similar findings. For example, Johnson (2005) and de Villiers and Johnson (2007) both tested children ranging from age 3 to 7 on their comprehension of tense/aspect and number agreement that is cued by third person singular -s. Across these studies they observed that AAE-speaking children did not comprehend this feature, while GAE-speakers began to show sensitivity to its grammatical interpretations approximately at age 5 to 6. These findings were later corroborated by Beyer and Hudson-Kam (2012) who assessed first and second grade AAE-speaking children on their comprehension of a variety of GAE morphological markers. They found that while GAE-speaking children demonstrated comprehension of some tense markers (i.e., past tense -ed, future contractable 'll, and third person -s), AAE-speaking children only showed accurate comprehension of morphological forms that overlapped between GAE and AAE (i.e., plural -s).

The prevailing explanation for the findings above have been entrenched in perceptual processing accounts of spoken language comprehension. A perceptual analysis account posits that linguistic differences between AAE and GAE create additional challenges in mapping a rapidly unfolding signal to underlying lexical representations (Floccia et al., 2006; Norris, 2003; van Heugten & Johnson, 2014). This explanation is not unlike previous hypotheses that emerged to explain connections between dialect and reading (i.e., the *linguistic interference hypothesis*). A perceptual processing account contrasts with the linguistic interference by emphasizing the relevance of the auditory signal during spoken language comprehension and children's knowledge and awareness of linguistic differences. This account has been useful for understanding how variable properties of the speech signal impact a listener's accuracy in accessing relevant word meaning, however, an inherent limitation of this explanation is that it evaluates linguistic processing in the absence of social parameters undergirding dialect use. Dialect variation is an inherently social behavior, an intricate web of the speaker's linguistic choices and the listeners' integration of the information provided both by the linguistic and non-linguistic properties of the speaker (i.e., speaker identity). Additionally, dialects are rooted in communities of talkers that vary systematically by any number of social factors. AAE is one example of this kind of variation because it is a raciolect, a dialect that is typically associated with African Americans. Moreover, children's early experiences with hearing or speaking AAE and GAE are associated with their lived experiences with other speakers of these varieties.

A growing segment of the literature has shown that listeners integrate nonlinguistic cues about the speaker (i.e., race and/or ethnicity) in ways that can either enhance or disrupt spoken language processing. Though the majority of this work has focused on adult listeners, the general trend in the literature is that listeners experience greater accuracy or ease of comprehension when the auditory signal coheres with their assumptions about the speaker (McGowan, 2015; Staum Casasanto, 2008). For example, studies have shown that adult listeners can understand American English with greater accuracy when the signal is paired with a photograph of a European American face, but with poorer accuracy when paired with a photograph of an Asian face (Babel & Russell, 2015; Kang & Rubin, 2009). Other studies have found that European Americans' comprehension of Mandarin-accented speech improves when the signal is paired with photographs of Asian speakers (McGowan, 2015). In these studies, listeners demonstrate a comprehension advantage in visual-auditory contexts when cues about the speaker's identity matches up to the listener's general assumptions both in terms of the auditory signal and the non-linguistic or visual aspects of the speaker 'providing' the signal. This interpretation aligns with evidence that Mandarin-accented speech improves in contexts when individuals observe faces of people most likely to produce that the signal or when American English is perceived as easier to comprehend when paired with faces of European Americans.

Listeners' assumptions may be a result of their lived and/or socialized experiences of different speaker communities, but the critical point here is that reasoning about the speaker is a fundamental aspect of spoken language processing. Moreover, nonlinguistic aspects of the speaker's identity (i.e., visual cues indexing race or ethnicity) and linguistic information provided through the speech stream can facilitate or inhibit the ability to reason about the mental and knowledge states of various interlocutors. One question that emerges from this literature is whether children are sensitive to both visual and auditory cues that guide linguistic processing. A few studies have shown that children are sensitive to racial cues fairly early in life (i.e., as early as 3 months of age) (Bar-Haim et al., 2006; Kelly et al., 2005; Wheeler et al., 2011). For example, some studies have shown that at 9 months of age children demonstrate a preference for faces that are familiar or similar to the child's race/ethnicity (Anzures et al., 2009; Kelly et al., 2009; Wheeler et al., 2011) relative to faces that are different in race.

To our knowledge, there have been three studies that examine children's integration of visual aspects of the speaker (i.e., race/ethnicity) during spoken language processing. Weatherhead and White (2018) investigated 16-month-old children's recognition of known words in both a familiar and less familiar accent when each of these signals were paired with either a person of similar or different racial background. Weatherhead et al. found that children fixated more to a target image in a familiar accent relative to an unfamiliar accent when infants saw a person of the same race, but looked equally to target images across familiar and unfamiliar accents when they saw a person of a different race. More recently, Erskine (2022b) examined European American and African American 6- to 11-year-olds recognition of words in semantically predictable sentences (i.e., sentences with predictive verbs) in three speaker conditions. These speaker groups included the following group membership conditions: a) dialect and race of speaker matched the listener, b) dialect and race of speaker differed from the listener, and c) either the dialect or race of the speaker matched the listener. Erskine et al. found preliminary evidence that access to visual and auditory speaker identity cues influenced children's overall word recognition and processing speed. That is, children looked more rapidly and efficiently to familiar words in conditions where speaker identity cues that were the most reliable for listeners resulted in rapid comprehension (i.e., dialect use for European American listeners and race- and dialect-related cues for African Americans). Lastly, Weatherhead et al. (2021) demonstrated evidence of word learning flexibility in the presence of visually salient socioindexical speaker cues (i.e., race). First, they showed that when monolingual or bilingual children learn words by speakers of the same race, each linguistic community demonstrated evidence of mutual exclusivity. This contrasts with some earlier findings in the literature which have suggested that bilingual children do not use mutual exclusivity in word learning to the same degree as monolingual children. Second, for conditions in which the speaker was of a different race, monolingual children relaxed their adherence to mutual exclusivity principles when learning new words. Each of these studies are interesting for several reasons. First, they provide converging evidence that children avail themselves to visual and auditory information about the speaker in ways that directly impact spoken language processing (i.e., comprehension or learning). Second, these studies suggest that the degree to which particular speaker identity cues (i.e., accent, dialect, race/ethnicity) impact real-time comprehension depend on the listener's lived experiences and the demands of the task.

Intriguingly, while two of the aforementioned studies (Erskine, 2022b; Weatherhead et al., 2018) find evidence that speaker identify influences comprehension in linguistic contexts where the demands to integrate speaker identity might be attenuated, Weatherhead et al. (2021)

shows evidence of speaker identity integration in a more socially influenced learning task where the demands to integrate speaker-related cues are even more paramount. For example, when listeners hear familiar words in isolation or in sentences such as *find/kick the ball*, the communicative intent does not lose its clarity. The sentence retains its semantic interpretation that there is an action that is being performed on an inanimate object (ball). However, mutual exclusivity, as tested in Weatherhead et al. (2021), requires listeners to generate an additional inference that a novel word refers to the referential candidate that is not known. The listener makes an assumption that if the speaker intended to refer to the known word, they would have used the word that is familiar to the listener. This pragmatic inference which underlies learning for many monolingual children and bilingual children, in some learning contexts, is based on the listener's reasoning about the learning environment and the likelihood that a certain speaker is referring to a novel object as opposed to the same object. Therefore, in contrast to lexical or sentential contexts where semantic interpretations are accurate (as in the example above), words or sentences that carry ***pragmatic interpretations*** require listeners to go beyond literal semantic interpretation (i.e., *explicit* interpretations) and retrieve the speaker's intended meaning (i.e., *implicit* meanings) (Grice, 1989; Huang & Snedeker, 2009; Papafragou & Musolino, 2003, *inter alia*). A learning task like the one used in Weatherhead et al. captures pragmatic reasoning in the case of learning, but there are also similar pragmatic interpretations that are derived during spoken language comprehension.

A classic linguistic phenomenon that entails this form of pragmatic reasoning are implicatures which are associated with scalar or logical terms like *some.* According to most theoretical accounts of scalar quantifiers, interpreting *some* necessitates that listeners suppress an explicit (semantic) interpretation that *some* could also mean *all* and compute the pragmatic inference that the speaker intends to restrict their interpretation to exclude an *all* interpretation. A central part of computing this inference is the assumption that if speakers produced *some*, then they do not intend to convey an *all* interpretation. Further, if the speaker intended to convey an *all* interpretation, then they would have said *all.* Take for example the following sentence:

(a) Billy has some of the balls

A purely explicit interpretation of sentence (a) indicates that Billy has *some*, and potentially *all* of the balls. However, a pragmatic interpretation, derived from counterintuitive reasoning would lead the listener to consider that if Billy had all of the balls, then the speaker would have used the maximum *all* and not a weaker logical term *some* (i.e., listeners grant preference to a subset interpretation and exclude the maximal interpretation). This level of pragmatic reasoning is a direct result of the listener's ability to derive the speaker's intent.

Psycholinguistic studies of scalar implicatures have primarily focused on the emergence of pragmatic interpretations among listeners. Using truth value judgment tasks, several studies have shown that while listeners can generate scalar inferences (i.e., a restricted interpretation of some but not all), this interpretation usually takes more time than accessing the literal interpretation. For example, Bott and Noveck (2004) found that when evaluating the veracity of the statement *Some of the elephants are mammals,* listeners who responded accurately (i.e., false) took more time than listeners who responded incorrectly (i.e., true). This evidence was taken to suggest that pragmatic inferences take time and are not immediately available to listeners.

Despite a preponderance of similar findings from earlier studies using truth value judgment tasks (Noveck & Posada, 2003; Rips, 1975), there has been some debate in the literature about the contexts that influence scalar implicature. To date, the understanding is that listeners' derivation of scalar implicatures is task dependent (Grodner et al., 2010; Noveck & Posada, 2003; Smith, 1980). Several recent studies have demonstrated that when the demands of the tasks are minimal, participants demonstrate increased accuracy and more rapid computation of scalar interpretations. For example, to evaluate the truth of the sentence *Some elephants are mammals*, listeners must reason about the potential interpretations for this statement as well as access their long-term knowledge about elephants and whether there are subsets of elephants that are not mammals. Given the demands of this type of task, it has been argued that delays in scalar implicatures computation are not surprising (Bott & Noveck, 2004; Noveck & Posada, 2003).

Simpler tasks that do not require these additional cognitive demands have provided evidence that adult listeners can quickly generate scalar implicatures. Using the visual world paradigm, Grodner et al. (2010) presented adult participants with sets of objects (e.g., balls and balloons) that were divided among four characters while they heard the one of three quantifiers within the phrase (*Click on the girl who has [some/none/all] of the balls).* In their study, they found that within 200ms of hearing the quantifier, individuals quickly and reliably fixated towards the subset interpretation. Similar speed of processing was observed by another group of adult participants in Huang and Snedeker (2018) which evaluated comprehension scalar quantifiers in different environmental conditions: a) conditions in which objects variably mapped onto exact numbers (two/three balls) or scalar terms or b) conditions in which objects were consistently described using scalar quantifiers (all/some). They found that when objects were consistently paired with a scalar label, as opposed to having a single object set (balls) mapped variably onto scalar and exact numbers, adults showed rapid comprehension of the scalar term *some*, almost as rapid as their comprehension of target words paired with the term *all.*

The developmental literature has also paid some attention to understanding how and when children become adult-like in their use and comprehension of scalar quantifiers. The existing body of literature mimics earlier findings in the adult literature that when provided with sufficient time (Huang & Snedeker, 2009) and the use of modified experimental tasks (Foppolo et al. 2012) children as early as 5 and 6 years of age can generate or comprehend scalar implicatures like *some*. For example, using the visual world paradigm, Huang and Snedeker (2009) examined 5- and 6-year-olds comprehension of the scalar terms *some* while viewing visual displays that contained referents corresponding to a subset or full set of objects (i.e., two of four socks or three of three soccer balls). Huang et al. examined children's selection of the subset of objects (socks) when it was referred to using either the scalar term *some* (i.e., Point to the girl with some of the socks) or an exact number (i.e., Point to the girl with two of the socks). Across trials in which the subset interpretation (i.e., displays referencing two of four socks) was mapped variably to either the scalar term *some* or the exact number term *two,* children demonstrated a delay in fixating to socks when the visual display was accompanied with critical sentences that contained the scalar quantifier (*some*). In fact, most children fixated to the subset interpretation only after hearing the noun referent (socks) unfold. These results indicate that while children can generate scalar inferences, they are drastically delayed in real time as compared to adult listeners. One critical limitation of this study was the variable use of labels for subset interpretations. In fact, in a later study by Huang and Snedeker (2018) it was acknowledged that potential delays in generating scalar implicatures may have been attributed to the task and the fact that two different terms (i.e., the logical scalar term *some* and an exact number *two*) were used to refer to objects corresponding to a subset interpretation. Therefore, a complex methodological design may have been responsible for delays in scalar inferencing among children. Similar evidence has also been found by Foppolo et al. (2012) who used sentence evaluation tasks to examine children's ability to generate scalar inferences. Using a classical truth value judgment task in which children made metalinguistic truth/false judgments for statements containing the scalar quantifier some, 4- and 5-year-old children demonstrated poor accuracy in generating the pragmatic inference. However, when the task was modified to provide children with additional metalinguisitc training about the under-informativity of the scalar term *some*, performance among 5-year-old children drastically improved.

Altogether, the research on scalar implicatures has shown that: a) comprehension of scalar interpretations (*some, but not all*) places greater demands on the listener's ability to reason about the speaker and their communicative intent and b) the complexity of the task and the degree to which metalinguistic judgments are required can influence children's overall ability to comprehend scalar implicatures. This study directly examines the effects of speaker identity and dialect variation on children's computation of scalar implicatures in an experimental task where children are more likely to be successful in deriving scalar implicatures. We focus on scalar implicatures because the majority of the research on spoken language comprehension involves literal semantic interpretation, but as noted above, a great deal of what speakers communicate is entrenched in implicit meanings instead of the literal meanings (Janssens & Schaeken, 2012; Grice, 1975). Additionally, scalar inferencing may require listeners to attend more to information about the speaker's use of language and non-linguistic aspects of the speaker's identity to derive an inference regarding the speaker's intended meaning (i.e., *some, but not all of the balls).*

To our knowledge, this is the first study to examine how socioindexical properties of the speaker impact children's comprehension of scalar implicatures. Using a modified adaptation of the visual world paradigm presented in Huang et al. (2009), we examined the impact of non-linguistic and linguistic speaker identity cues on children's comprehension of the scalar term *some*. During this eye-tracking task, children heard stories that corresponded to a visual display containing two object sets that were distributed among four separate characters. Gaze patterns to the visual display were monitored as children heard sentences that contained scalar quantifiers (some/all).

In two linguistic communities, European American monodialectal GAE listeners and African American listeners who vary in their use of AAE and GAE, two research questions were addressed: (1) Do visual and auditory speaker identity cues influence pragmatic inferencing in scalar implicature; and (2) How are visual and auditory cues weighed by children as they compute pragmatic interpretations across three different speaker group conditions: a) seeing an African American speaker and hearing AAE, b) seeing an African American speaker and hearing GAE, and c) seeing a European American speaker and hearing GAE?

Experiment 4.1: European American Listeners

Experiment 4.1 examines the two research questions above in a group of 6- to 11-year-old European American listeners. Our predictions are based on the integration of two findings in the literature, the previous Erskine (2022b) study examining the effect of speaker identity and dialect variation on semantic prediction and Weatherhead et al. (2021) who examined race effects of mutual exclusivity among monolingual and bilingual children.

Weatherhead et al. (2021) showed that monolingual children who were primarily European American were more likely to demonstrate mutual exclusivity when learning words from speakers of a familiar race. By contrast, mutual exclusivity was relaxed when they learned words from a speaker of African descent (i.e., African American). This differs from the results observed by Erskine (2022b) who found that children showed a dialect preference in which they efficiently processed words by speakers that shared the same dialect regardless of race. In the current study, two outcomes are possible. It may be that children will demonstrate a race-based distinction in which they are able to compute scalar implicatures with greater efficiency from speakers of the same race (i.e., seeing a European American and hearing GAE), aligning with the results reported by Weatherhead et al. They will be least efficient in making scalar predictions when they see an African American speaker regardless of the dialect spoken (GAE or AAE). Alternatively, children may continue to demonstrate a dialect-based preference that is even more pronounced because the comprehension task is more difficult. Specifically, they will demonstrate efficient scalar inferencing for all speakers of GAE regardless of race (i.e., European American or African American speakers of GAE), and less efficient scalar inferencing for speakers of AAE.

Methods

Listeners

Twenty European American children (13 female) between the ages of 6;0 and 10;11 (years;months) (mean=8;8, sd=1;6) were included in this study. Children from areas near Washington, D.C., Maryland and Virginia were invited to participate during the March-to-August 2022 period of the pandemic. An additional 5 children were recruited but not included in the final sample for the following reasons: a) equipment failure or missing data that exceeded the threshold for inclusion (n = 3) and b) failure to complete all of the language measures (n=2).

Children were primarily recruited from a University-based consortium of families who expressed interest in participating in child development research. Families from this database were provided information about the study and invited to participate through email. Other families who participated heard about this study through word-of-mouth from families who had previously participated; these children usually attended the same school as previous participants or lived in the same neighborhood (n=5). The majority of participants included in this study were from families with high levels of maternal education (e.g., received a college degree or more), based on a demographic questionnaire filled out by parents. See Table 1 for further details.

Table 1
Demographic summary of listeners and their scores on standardized assessments in Experiment 4.1

n	20
Female	13
Mean Age (in months)	102 (19) *range=72-131*
PVT NIH Toolbox *Mean Raw Score* (SD)	80 (7) *range=67-94*
Mean Standard Score (SD)	111 (10) *range=91-131*
DELV-ST *Dialect Density Score*	.07 (.08) *range=.00-.27*
Maternal Education Levels	
Low (less than high school)	1
High (college degree or higher)	19

Experimental procedures

Visual world paradigm. All children completed in-person research visits at the University. A within-subjects design was used for the eye-tracking study. Sentences across all three speaker group conditions were pseudorandomized offline and then administered in a single block of eye-tracking. Gaze patterns were monitored using either an automatic SR Research Arm-mounted Eyelink 1000 Plus (n = 13) or a portable Duo Eye-tracking System (n = 7). The experiment was designed to model a space galaxy quest game in which children were rewarded with a missing "item" or "crew mate" while participating in the experimental task. During the course of the experiment, trials were grouped into 6 smaller chunks for administration purposes, separated by a visual reward (i.e., a picture of missing "item" in their quest) to maintain children's participation throughout the task and to provide breaks when required.

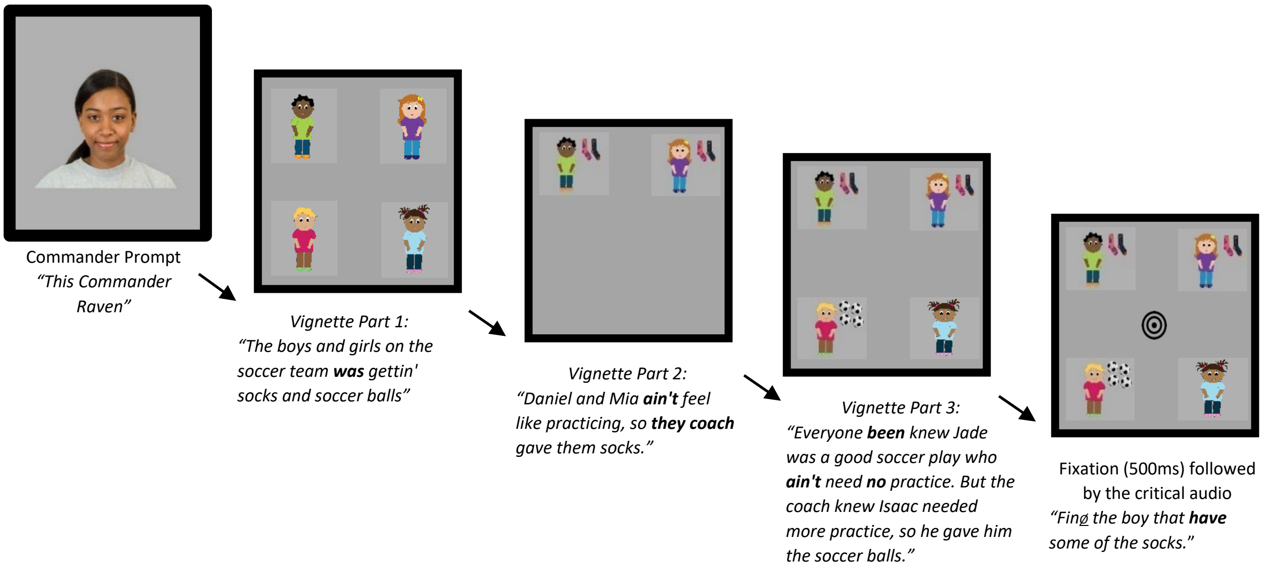
At the start of the experiment, children were informed that they would be introduced to specific commanders who would tell them a story about two girls and two boys. Children were instructed to listen carefully to each story and to select the image belonging to the character that best-matched the sentence that was heard using an electronic touchscreen stylus. After the initial instructions, children were introduced to all three commanders one at a time, and were presented with an auditory introduction from each commander that was recorded in the dialect assigned to each speaker group condition (*"Hi. My name is Commander [name]. I'm going to tell you a story about a group of boys and girls. Listen carefully!"*)[4]. This procedure was used to increase children's visual attention to the speaker images and auditory stimulus throughout the course of the experiment.

[4] The introductions in AAE included one linguistic feature difference, *going to* was produced as *gon* in AAE.

Trial event sequence. At the beginning of each trial, children were shown an image of one of the commanders while they heard a short prompt that reintroduced the Speaker (*"This is Commander [name]*." After 500ms, the image of the speaker disappeared and a screen showing all four characters without the noun referents was displayed. The vignettes that were separated into 3 audio files were played one a time to contextualize the arrangement of images on the screen and advanced in the following order: a) during the character-only screen children heard the first part of the story *"The boys and girls on the soccer team were getting socks and soccer balls on the field",* b) images assigned to the top-left and top-right were shown while children heard the audio, "*Daniel and Mia did not feel like practicing, so their coach gave them socks",* and c) while the top images remained on the screen, images assigned to the lower quadrants were displayed while children heard the remainder of the story, *"Everyone already knew that Jade was a good soccer player who didn't need any practice, but the coach knew Isaac needed more practice so he gave him the soccer balls."*

Vignettes and images were presented in this order to reduce memory demands and provide additional context about the distribution of the images on the screen. After hearing the vignette, the 4 images remained on the screen and a fixation attention getting video (e.g., a bullseye) appeared in the center of the screen. Children's fixation to the attention getter for at least 500ms initiated the onset of the critical auditory stimulus (some/all sentences). If the children required additional prompting, the experimenter would remind children to listen to the entire sentence and then select the picture that matched what they heard. The trial terminated after children selected an image using the stylus, which then initiated the start of the next trial.

Figure 1. Sequence of an example trial (*condition*: African American Speaker/Heard AAE).



Language measures. In addition to the visual world paradigm, children were administered two assessments, one of receptive vocabulary and one that evaluated children's use of U.S. English dialect patterns. The measure of receptive vocabulary was taken from the NIH Toolbox Cognition Battery which is appropriate for children between the ages of 3 and 85 years old. Additional details of the development and assessment and scoring procedures can be found in (Weintraub et al., 2013). Participants were shown a 2x2 array of images on an iPAD and heard an audio recording of a word. Children were provided with instructions to select the image that best represented the meaning of the word that was heard. The receptive vocabulary measure uses an adaptive testing procedure in which the number of items administered depend on children's accuracy of the previous word. Children are administered items one at a time until the task is completed. This measure provides a variety of scores based on item response theory. In the statistical analyses, we used the "uncorrected standard score" computed directly from the NIH toolbox. While the manual labels it as a type of "standard score" (Weintraub et al., 2013), in actuality it is a raw score which evaluates the participant's performance or vocabulary knowledge relative to other children in the U.S. In this paper we use the descriptive the label provided by the Toolbox but acknowledge that this score does not refer to conventional standard scores. We also administered the DELV Screening Test Part I (Seymour et al., 2003) to quantify and categorize children's dialectal patterns. We included this measure as a way to confirm the primary dialect of children. Children's range of dialect density scores (calculated as the number of NMAE dialect features by the total number of GAE and NMAE responses as in NP Terry et al., 2010) were very low so we did not include this measure in any of the statistical analyses.

Eye-tracking Materials and Stimuli

Vignettes

As described above, all test sentences were preceded by a story vignette. Each vignette (a) was included to contextualize the arrangement and distribution of images across the four quadrants of the display screen.

(a) "The boys and girls on the soccer team were getting socks and soccer balls on the field. Daniel and Mia did not feel like practicing, so their coach gave them socks. Everyone already knew that Jade was a good soccer player who didn't need any practice, but the coach knew Isaac needed more practice so he gave him the soccer balls".

Sentences

The experimental task consisted of a total of 27 sentences. Of these sentences, 12 sentences contained the quantifier *SOME* and the remaining 15 sentences were fillers trials that contained the quantifier *ALL*. The 27 sentences were distributed across three speaker group conditions: a) *Same Race|Heard GAE*; seeing a European American and hearing GAE, b) *Different Race|Heard AAE*; seeing an African American and hearing AAE, and c) *Different Race|Heard GAE*; seeing an African American and hearing GAE. Children heard 9 sentences in each of the speaker group conditions: 4 SOME trials (*Find the girl that has some of the socks*) and 5 filler trials (*Find the girl that has all of the soccer balls).* A complete list of the experimental trials can be found in the Appendix (Table H1 and H2). To reduce the complexity of the task and ensure that all children would generate the scalar inference, the critical sentences contrasted scalar quantifiers *some* and *all*, and did not include other logical terms (e.g., *none* as has been found in previous studies such as Huang & Snedeker, 2009, 2018; Grodner et al. 2010).

AAE features included in the stimuli

In each sentence and vignette, phonological and morphosyntactic differences between AAE and GAE were included. In the critical sentences, two AAE features were included: a) zero-marking final consonants in consonant cluster and b) use of (***have)*** for the present perfect tense verb (has) in GAE. These linguistic differences resulted in the following AAE sentence: "*Fin<u>ø</u> the girl that have some of the socks.*

Among the vignettes the following AAE features were included: a) subject-verb agreement (was/were), b) negative contraction (ain't), c) alveolarization, and d) stressed BIN, e) possess they, and f) negative concord. The inclusion of these AAE features resulted in an example AAE vignette below (b).

(b) "The boys and girls on the soccer team **was** gettin' socks and soccer balls. Daniel and Mia **ain't** feel like practicing, so **they coach** gave them socks. Everyone **been** knew Jade was a good soccer play who **ain't** need **no** practice. But the coach knew Isaac needed more practice, so he gave him the soccer balls. "

Visual displays

The visual displays included images of the objects and the four characters that were labeled and described in the vignettes above. Each cartoon character was assigned to one of the four quadrants in the following positions: Daniel (top-left), Mia (top-right), Isaac (bottom-left) and Jade (bottom-right). The arrangement of characters in each quadrant remained fixed across all trials, this was done to decrease task-related memory demands and to familiarize children with the location of each cartoon character on the screen across trials. Vertically-aligned quadrants contained characters that matched in sex (male-male or female-female) and horizontally-aligned quadrants consisted of characters that differed in sex (male-female).

The two sets of objects that were mentioned in the vignette (e.g., socks and soccer balls) were distributed among the four characters in the following ways: two characters were given a subset of one of the items (two socks each), one character received none of the items, and the remaining character received the full set of the other object (four soccer balls). Figure 1 shows an example of the display screen for the trial (socks/soccer balls).

Further, the object sets in all of the *SOME* trials and 6 of the *ALL*-filler trials contained objects that shared a phonological onset (i.e., *socks* and *soccer balls*). Phonological overlap across two of the items created a brief period of ambiguity between the target noun referent and one of the competitor items on the screen. The remaining filler trials (n=9) contained object sets that were phonologically unrelated (i.e., *apples* and *carrots*).

Fourteen stimuli items were taken from Huang et al. (2009). An addition 13 items were included to increase the number of trials and to ensure that words retained temporary phonological ambiguity at the onset of words across the two dialects that were included in this study (AAE and GAE).

Stimuli processing

Auditory stimuli. Sentences were recorded by three female 21- to 25-year-olds who self-identified as belonging to one of the following ethnicity and linguistic communities: a) a European American monodialectal speaker of GAE, b) an African American monodialectal speaker of GAE, and c) an African American bidialectal speaker of AAE and GAE.

All auditory stimuli were recorded using a Shure SM51 microphone in a sound-attenuated booth. The sentences were normalized for duration across dialect conditions through Praat (Boersma & Weenink, 2022) by either lengthening or shortening the duration of the segment normalized. The length of the quantifier sentences were 2603ms. Stimuli were time locked across three regions relative to the onset of the sentence: a) 450ms, which is the onset of the gender region (*girl/boy that has*), b) 1000ms, which is the onset of the quantifier (some/all of the) (henceforth, *the quantifier region*), and c) 1600ms, which is the onset of the target referent (socks).

The vignettes were not time locked and varied in length, ranging from 8.5 seconds to 12.3 seconds (mean = 9.8 seconds, sd = 1.2 seconds). The vignettes were separated into three separate audio files to be presented one at a time to accompany the distribution of items across the four quadrants.

All sentences and vignettes were normalized to the sound level of 70 dB SPL.

Images

Noun referents. All words were pictured with images or available from public online databases. One image of each referent was selected because most of the trials contained unique referent object sets. However, for the filler trials, 2 of the object sets were repeated because we needed to distributed 4 unique homophonous filler object sets (i.e., laptops and ladders) across three speaker group conditions. To do this, two object sets were randomly selected and were shown twice, once in one speaker group condition and again in a different speaker group condition.

All noun referent images were edited using Adobe Acrobat Photoshop so that all objects were approximately the same size, 150 pixels on the longest side. The edited objects were then superimposed on a 600x600 pixels canvas that contained one of the cartoon characters described below. This experiment was designed during the height of the pandemic, so the images were not normed prior to administering the experiment. However, we included objects with an younger age of acquisition than the children recruited in this study and the vignettes which accompanied the display of images in a sequential order provided additional opportunities for children to become familiar with the object labels.

Cartoon characters. Illustrations of the cartoon characters were commissioned from a local artist in Maryland. All illustrations were originally designed in Adobe Illustrator and then edited in Adobe Photoshop (Adobe Inc., Version CC 2019). The dimensions (width and height) of characters were adjusted so that they could fit on a 600x600 canvas and were projected onto a grey background. As mentioned earlier, images of noun referents were also added to displays to capture the distribution of items across the characters. The cartoon characters included pictures of two boy and girls of diverse cultural backgrounds to increase the representation of different backgrounds of children in the study. The images of children did not change over the course of the experiment because we wanted to reduce attentional demands and highlight the distribution of object sets as opposed to the characters themselves.

Speaker images. To evaluate the contribution of nonlinguistic speaker cues, *race/ethnicity*, photographs of each "speaker" were included in this study. Photographs were taken from the Chicago Face Database (Ma et al., 2020), a repository of photographs of 597 unique individuals. This database contains faces of individuals who self-identified as males or females from the following race and ethnicity backgrounds: Asian, African American, LatinX, and European American. Extensive norming procedures were undertaken for each individual photograph. Subjective perceptual ratings of socially and psychologically meaningful attributes (e.g. attractiveness, friendliness, baby-facedness, and several others) and objective measures of physical attributes (i.e., jaw measurements, angularity of the face, etc) were obtained during norming procedures. Ratings were acquired from a sample of 1,087 individuals from diverse cultural and linguistic backgrounds. Additional details, normative data, and procedures can be found in the publication of the database (Ma et al., 2015).

For the purpose of this experiment, we initially selected eight images, four unique pictures of European American or African American females with happy, closed-mouth expressions to ensure an implied friendly demeanor. The advantage of using this database is that photographs of each individual exhibited uniform poses, attire, and expression. In other words, all individuals wore the same item of clothing, had similar facial expressions (e.g., a close-mouthed smile) and posed facing forward.

A growing body of research (Blair et al., 2010; Zebrowitz & Montepare, 2008; Zebrowitz, Voinescu, & Collins, 1996) has shown that less desirable social attributes are typically associated with individuals from minoritized backgrounds on the basis of visual cues. To minimize the extent to which socially-biased inferences would interfere with the experimental task, an informal panel of adults from diverse cultural and linguistic backgrounds was convened to select images that activated more neutral assumptions about the speaker's social prestige. This process resulted in the final selection of 4 European American photographs and 3 African American photographs. All images were edited using a commercial picture editing software Adobe Photoshop (Adobe Inc., Version CC 2019). Speaker images were cropped to 500x500 pixels and superimposed on a canvas of 600x600 pixels with a grey background.

Four experimental lists (each containing the 27 test items) were created based on the permutations of the selected images of Speakers across the three speaker group conditions. Because the number of photographs of African American and European American was unbalanced, in each list, one of the four European American images was randomly selected without replacement and assigned to the condition in which listeners saw a European American photograph and heard GAE. Additionally, in each list, two of three African American photographs were pseudorandomized and each photograph was then assigned to one of the two remaining speaker group conditions. This permitted the creation of lists that contained a unique photograph-speaker group condition pairing. Additionally, the image that was randomly assigned to a specific speaker group condition remained fixed over the duration of the entire experiment. This was crucial to ensure that participants had consistent opportunities to track the visual and auditory information provided by each speaker. Table H3 in the Appendix H contains a table showing image-condition pairings across each of the four lists.

Results

We report three separate analyses below for children's response patterns to *SOME* trials. Although this task included a few ALL trials, they were only included as fillers. Moreover, there were not enough ALL trials to include in the analysis. The first analysis evaluates the effect of speaker identity cues on children's *accuracy* in comprehending scalar implicatures. The second analysis uses the cluster-based permutation method to identify a region of analysis for examining speaker identity effects during real-time comprehension of scalar quantifiers (i.e., gaze patterns). The last analysis investigates the effect of dialect variation on scalar implicatures with additional predictors of language ability and age in the region of time identified by the cluster permutation method.

Response accuracy. Overall accuracy in this task was above 96% across all experimental conditions. A generalized linear model with response accuracy coded as 0 or 1 was used to examine whether scalar implicatures accuracy was influenced Speaker Group Condition (as a dummy-coded variable), Vocabulary Size (as a centered-raw score from the PVT) or Age (as a centered continuous predictor). None of the main effects were significant predictors of response accuracy. Incorrect trials (n=6) were excluded and all further analysis of the gaze data was restricted to accurate responses.

Gaze data preparation and cleaning. The data were carefully screened for missing data and cleaned using procedures provided by customized R scripts and the EyetrackingR. First, a *deblinking* procedure that interpolated up to 150ms of missing data was implemented (Github Script, Mahr, 2017). This procedure eliminated sources of missing data that were attributed to blinks. Next, the quality of data was examined in a window that began 250ms after the onset of the baseline region (*find the*) to 1500ms after the onset of the labeled referent (*socks*). This general window was selected to account for the time it takes children to program a saccade.

Children were excluded from all further analyses if they did not have at least 60% reliable data across one block of eye-tracking. We used a 60% criterion rather than a more stringent criterion because of challenges with the eye-tracking equipment that affected calibration for at least 6 participants.

These criteria resulted in the inclusion of 199 of 294 trials across participants. Ten additional participants were tested, but were excluded because they had greater than 60% of missing data. Among the remaining participants (n=20), the proportion of missing data did not differ across Speaker Group, $F(2, 196) = 1.87$, $p = .16$. Table 2 and 3 provide a summary of the track loss prior and after trial or participant exclusion.

Table 2.
Summary of the track loss and sample analysis before trial-exclusion (n=25).

Speaker Group	Number of Trials	Mean Samples	Percent Track Loss (%)
Saw AA Speaker: Heard GAE	99	2172	46.2
Saw AA Speaker: Heard AAE	97	2180	50.4
Saw EA Speaker: Heard GAE	98	2250	49.1

Table 3.
Summary of the track loss and sample analysis after trial-exclusion (n=22)[5]

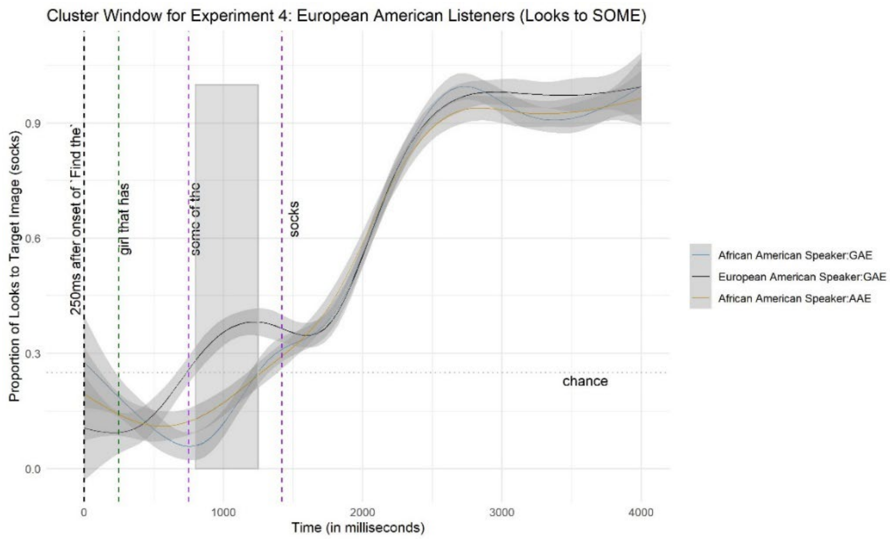
Speaker Group	Number of Trials	Mean Samples	Percent Track Loss (%)
Saw AA Speaker: Heard GAE	70	2095	38.6
Saw AA Speaker: Heard AAE	65	2061	40.7
Saw EA Speaker: Heard GAE	64	2000	37.5

Cluster Analysis. Previous studies on children's comprehension of scalar implicatures (Huang & Snedeker, 2009, 2018; Tavano & Kaiser, 2010; Yoon et al., 2015) had clear *apriori* predictions about when gaze patterns would depart from chance or some baseline of looks from the competitor image. However, these studies have examined the comprehension of scalar quantifiers in children's native dialects or languages. Moreover, these studies have only investigated children's attention to the auditory signal provided by the speaker. To our knowledge, this is the first study to evaluate the effect of speaker identity (i.e., as determined by both visual and auditory information) on scalar implicatures. For this reason, we employ the use of an objective approach (i.e., cluster analysis) to identity a region of time that would be most relevant to our experimental manipulations. Further, this approach minimizes the risk of Type I error by allowing us to examine a single region of time as opposed to examining effects across multiple regions of time. We fit a mixed effects model predicting the log odds of looking to the target image (the subset interpretation) which included a fixed effect for Speaker Group, a by-subject random intercept and slope for Speaker Group, and a by-item random intercept. The

[5] Two participants were excluded from the final gaze analysis (the generalized linear mixed effects model) predicting the likelihood with which participants looked to the target image (socks) because they did not complete the language measures which were included as covariates in the model.

cluster analysis identified a cluster of 10 contiguous 50ms bins that began 50ms after the onset of the quantifier (some) until 100ms prior to the onset of the noun referent (socks) (see Figure 2).

Figure 2. Looks to the target image (socks) when hearing SOME trials in the cluster analysis region.



Generalized linear mixed effects model. A generalized linear mixed effects model was fit to the cluster window to examine the influence of the three speaker group conditions and additional child-related measures of language and linguistic maturity on their fixation to the correct interpretation (i.e., the image corresponding to the subset of referents). The dependent variable was the log odds of looking at the target image. The predictor variables included fixed effects for Speaker Group (a contrast coded categorical variable), Vocabulary (as a centered continuous raw score on the PVT), and Age (as a centered continuous variable). Additionally, the model included two-way interactions between Speaker Group and all other aforementioned independent variables.

Contrast coding was used to evaluate the prediction that European American listeners would prioritize speaker identity cues based on the speaker's language use above and beyond cues related to the speaker's race. Therefore, the model evaluates two specific contrasts of

Speaker Group. In each contrast the reference level corresponds to the speaker group condition, *Same Race|Heard GAE*, (i.e., hearing GAE and seeing a European American). The first contrast corresponds to differences in children's looks to the target image when seeing a European American speaker of GAE speaker or an African American Speaker of AAE, while the second corresponds to differences in between hearing speakers of GAE that differed in race (i.e., European American and African American speakers). The intercept in this model represents the log odds of looking at the target image averaged over the three speaker group conditions for individuals of average age and vocabulary size. The R syntax is provided in the Supplementary information.

A mixed effects logistic regression was fit across the region identified by the cluster analysis (1050ms-1500ms, corresponding to the onset of the quantifier to 100ms before the noun). We found a significant main effect of the second contrast of Speaker Group. This result indicates that children looked less to the target image when they saw/heard an African American GAE speaker. Lastly, we observed a marginal trend between Vocabulary and the first contrast of Speaker Group, indicating that children with larger vocabularies demonstrated a larger difference between European American Speakers of GAE and African American Speakers of AAE than children with smaller vocabularies. All other main and interaction effects were not significant.

Figure 3. Looks to the target image across experimental condition (*speaker group conditions)*.

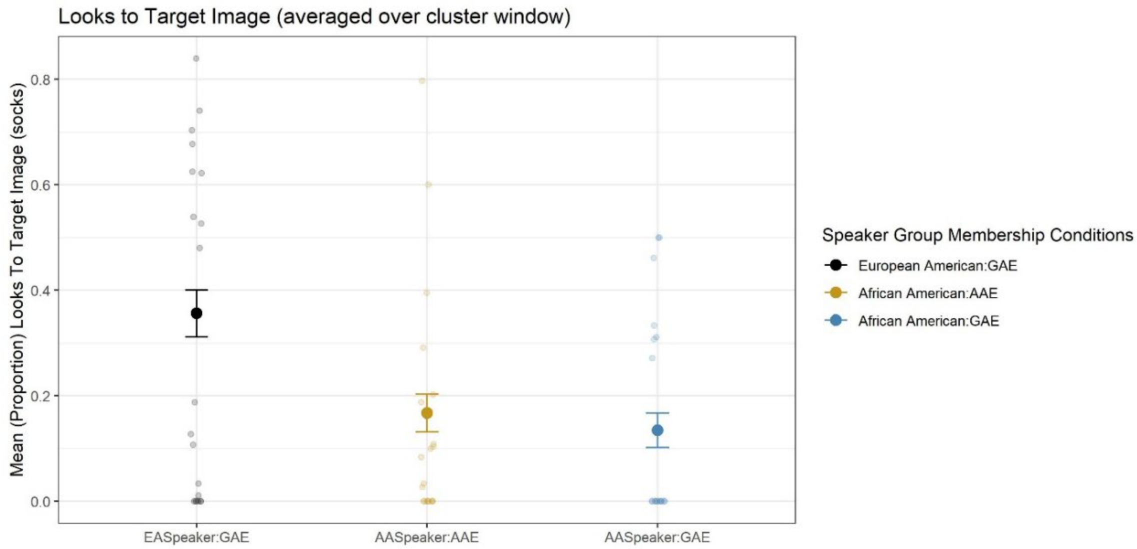


Figure 4. Looks to target image as a function of vocabulary size (*PVT NIH Toolbox*; median split), fitted as continuous in the regression analyses.

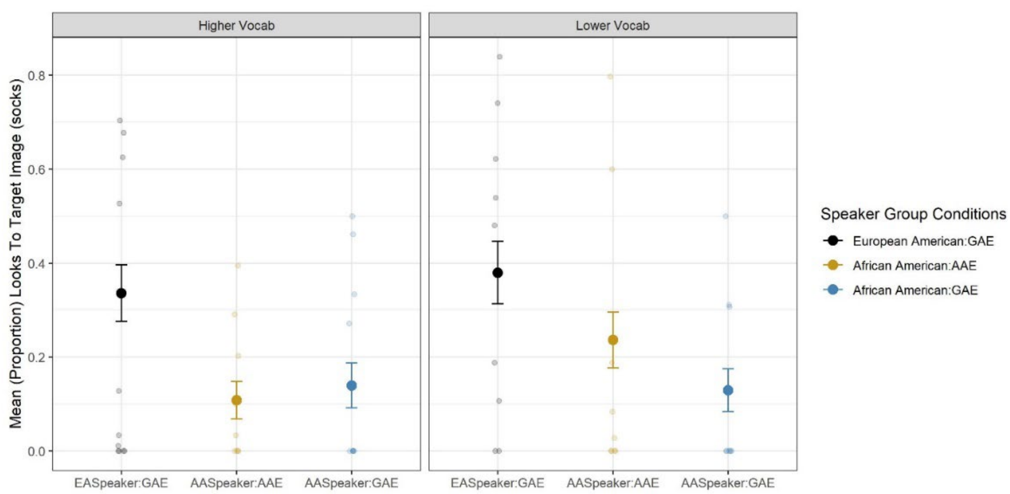


Table 4.
Summary of parameter estimates for the mixed-effects logistic regression model fit to the gaze. Model was fit using maximium likelihood.

R Syntax: glmer (cbind(LooksTarg, LooksNotTarg) ~ Vocabulary (*centered raw score*) + Age (*centered*) + Speaker Group (*contrast coded*) + Speaker Group *Vocabulary + Speaker Group *Age + (Speaker Group | participant) + (1 | trial)

Fixed Effects				
	b	**SE**	**z**	
Constant	-7.52	1.48	-5.06***	
Vocabulary (NIH Toolbox PVT Raw Score)	-.63	1.38	-.46	
Age (In Months)	.35	1.43	.24	
Speaker Group (Contrast 1)	-1.55	1.34	-1.16	
Speaker Group (Contrast 2)	-9.77	2.57	-3.80***	
Vocabulary*Speaker Group (Contrast 1)	3.26	1.81	1.81+	
Vocabulary*Speaker Group (Contrast 2)	1.71	2.44	.70	
Age*Speaker Group (Contrast 1)	-2.68	1.75	-1.53	
Age*Speaker Group (Contrast 2)	-.09	2.68	-.03	
Random Effects				
Parameter	**Variance**	**SD**	**Correlations**	
Participant-level				
Constant	32.73	5.72		
Speaker Group (Contrast 1)	26.55	5.15	.22	
Speaker Group (Contrast 2)	187.029	13.68	.80	.29
Trial-level				
Constant	9.78	3.13		

p <.001***, *p* < .01 **, *p* < .05*, *p* ≥ .05 +

Discussion

Previous studies have demonstrated that children are sensitive to socioindexical cues (linguistic and nonlinguistic) during the comprehension of sentences that contain explicit semantic interpretations (e.g., recognizing familiar words in literal semantic contexts, *find the* *ball*) (Erskine, 2022b). These studies suggest that listeners, including children, readily integrate information about the speaker in meaningful ways that impact the ease of spoken language processing. We sought to push the boundary of these findings by examining children's sensitivity to these same cues during pragmatic inferencing, these being sentential contexts that require children to integrate information about the speaker to retrieve implicit interpretations of sentences (i.e., deriving scalar implicatures).

This study provides additional evidence that when the demands of the scalar implicatures experiments are minimized (i.e., there are fewer ways to interpret a single object set in a visual display, see Breheny et al., 2013, Grodner et al., 2010, and Huang & Snedeker, 2018 for further discussion), children between 6 and 11 years old rapidly generate pragmatic predictions (i.e., as soon as the quantifier begins to unfold in the sentence). More crucially, this study found that children's successful scalar predictions were influenced by cues to the speaker's race. During the quantifier region of sentences in Experiment 4.1, European American Listeners exhibited a greater number of looks to the target image for contexts in which they saw a European American speak and heard GAE. However, looks to *some* remained close to chance for conditions featuring African American speakers regardless of the dialect that was heard. These results differ from the semantic prediction study (Erskine, 2022b) that found that children prioritize dialect over race. Here, we see the opposite trend in which European American listeners prioritized race over dialect, similar to Weatherhead et al. (2021). This is an interesting difference to observe in this study because it suggests that pragmatic inferencing tasks operate differently from other kinds of comprehension tasks that involve literal semantic interpretations. Specifically, it suggests that the ease with which listeners can represent another speaker's intent might depend on the alignment of several indices of the speaker's identity, at least in the case of implicit pragmatic interpretations. It appears that dialect alone may not provide sufficient information to facilitate the listener's ability to reason about speakers and their communicative intents.

In addition to the central research question of interest, we found an interaction between vocabulary and speaker group conditions representing either *Same Race|Heard GAE* or *Different Race|Heard AAE*. Children with smaller vocabularies, relative to children with larger vocabularies, looked more to the target image for speaker group conditions in which they saw an African American Speaker and heard AAE. Given the limited research on interactions among vocabulary size, pragmatic inferencing and speaker identity, it is unclear how to interpret this finding. This result raises interesting questions for future research, but cannot be easily explained by this study's design or the manipulations included in the eye-tracking task.

Experiment 4.2: African American Listeners

Methods

Experiment 4.2 examines listeners' comprehension of scalar implicatures while integrating visual cues to the speaker's race/ethnicity and auditory cues to the speaker's dialect among African American listeners with variable use of AAE or GAE. In this study, our expectations are based on findings from Weatherhead et al. (2021) in which bilingual children showed increased use of mutual exclusivity for word learning for speakers of the same race, but decreased adherence to mutual exclusivity for speakers of a different or less familiar race. We hypothesize that bidialectal speakers are more similar to bilingual speakers than to monolingual speakers in that they experience different linguistic patterns that are used to convey meaning in similar ways. Accordingly, we predict that the African American children will show more efficient pragmatic inferencing for speakers of the same race regardless of dialect, but will demonstrate greater uncertainty for speakers of a different race. These predictions also align with previous trends reported by Erskine (2022b) in which African American listeners rapidly and efficiently recognized words in either AAE or GAE that were paired with African American faces, but showed the least efficient recognition of words when words were produced in GAE and paired with a European American face.

Participants

Listeners

Twenty African American listeners (14 female) between the ages 6;0 to 11;0 (years;months) (mean=8;4, sd=1;7) were included in this study. Children were recruited during the height of the Omicron wave of COVID-19 in the United States. To ensure an adequate sample size, the geographic reach was widened to include children from states on the broader eastern coast of the United States. As a result, we included children from areas near New York, Maryland, Virginia, and Washington D.C. Eleven additional children were recruited but were not included in the final sample for the following reasons: a) children did not complete the eye-tracking task due to poor compliance or equipment failure (n =9) or b) children did not complete the norm-referenced language assessments (n=2).

Children were recruited from three different places: a) the University-based consortium of families (n=11), b) a local Maryland chapter of the National Association for the Advancement of Colored People (NAACP-MD) (n=5) and c) a local church in New York (n = 4). The majority of participants included in this study were from families with higher levels of maternal education (e.g., received a college degree or more, n=17). To accommodate flexible participation in research, we tested children either virtually or in person using webcam eye-tracking (see VWP Procedures for further details). Although we included remote participants in our analyses, we had a very small sample of children who received this method (n=2) and were included in the study. A total of four children initially participated in this study remotely, however two of these participants had a large amount of missing data due to attentional, equipment, or compliance difficulties. These challenges are inherent to remote eye-tracking research where there is less control over the testing environment and more opportunities for distractible environments.

The final sample included children with a history of typical speech or language development and a range of dialect density scores on the DELV-ST. Twelve children were categorized as speakers of GAE (*No* Variation), four children demonstrated *Some* Variation and four children demonstrated *Strong* Variation. See Table 5 for additional demographic information.

Table 5.
Demographic summary of listeners and their scores on standardized assessments in Experiment 4.2

n	20
Female	14
Mean Age (in months)	100 (17) *range=72-132*
PVT NIH Toolbox *Mean Raw Score* (SD)	79 (8) *range=60-93*
Mean Standard Score (SD)	105 (15) *range=72-131*
DELV-ST *Dialect Density Score*	.23 (.18) *range=.00-.93*
Maternal Education Levels	
Mid (technical/associate's degree)	3
High (college degree or higher)	17

Visit Procedures

The experimental procedures described in Experiment 4.1 were the same as for Experiment 4.2 with two exceptions: a) the Diagnostic Evaluation of Language Variation

(*DELV-ST*, Seymour et al., 2003) was used to confirm children's linguistic repertoire and to obtain a measure of dialect density[6] and b) eye-tracking data was collected through webcam recordings as opposed to automatic eye-tracking. These differences were a result of a change in data collection procedures during the pandemic. Data collection for African American listeners began prior to data collection for European American listeners. Further, data collection for African American listeners began at the height of the Delta/Omicron waves. Because families were reluctant to come into the lab for in-person visits, we decided to use online webcam eye-tracking. Ultimately, because of concerns about data quality, it was necessary to transition to in-person visits with African American participants. We continued to collect data using webcam procedures so that we could compare patterns across children who participated in either remote visits or in person.

Webcam eye-tracking. Previous studies have shown that webcam eyetracking is a useful tool that can be used to evaluate linguistic phenomena among populations typically excluded from laboratory-centered research. Moreover, studies have shown webcam eyetracking is able to replicate results that have been previously reported from studies in more controlled laboratory settings (Semmelmann & Weigelt, 2018; Venker et al., 2020; Yang & Krajbich, 2020). Nevertheless, webcam eye-tracking introduces additional sources of variability that can impact gaze data quality (i.e., challenges related to screen size and camera quality or increased opportunities for environmental distractions). To minimize these challenges, some precautions were taken. For example, participants were asked to sit in a quiet room with an adult nearby to

[6]A proportional score that measures children's use of AAE or nonmainstream American English features. Scores range from 0 to 1, scores closer to 1 indicate more AAE usage and scores closer to 0 indicates less AAE usage. The score is derived by taking the total number of Column A responses on the DELV and dividing it by the sum of Columns A and B. We included a dialect density score in subsequent analyses to capture children's variable use of AAE and GAE. Dialect density was calculated as a proportional score of columns A and B across the 15 test items. This approach is commonly used in the dialect research literature to quantify dialect use among African American students (Craig et al. 2014; Terry et al. 2006, Terry et al. 2010; Terry et al. 2012; Washington et al. 2018).

support the remediation of technical challenges and the four images on the screen occupied coordinate positions that placed them close to each of the corners on the screen. These decisions were necessary to reduce the challenges introduced by these sources of variability and assist with later manual coding procedures. The visual world paradigm was implemented using Penn Controller for Ibex (Zehr & Schwarz, 2018). A combination of customized java scripting and default functions provided by Ibex were used to create the trial sequence that was described in Experiment 4.1. The experiment began with a welcome page that included an embedded consent form and instructions on how to set up the webcam. After completing informed consent, caregivers gave the browser permission to use the webcam which recorded children during the course of the experiment.

Once the camera was set up, children were seated in front of the computer screen at a relatively close distance to the child for comfortable viewing of the images. They were then provided with general instructions about how to complete the Galaxy Quest game. Prior to meeting the commanders and completing text trials, children completed eight practice trials to help familiarize them with the experiment and provide opportunities for manual eye-tracking coders to later determine whether gaze patterns were mirrored during the recordings. Additionally, because we needed to obtain accuracy during remote visits, verbal responses were collected instead of a touch response. To evaluate accuracy, pictures appeared on the screen with a border of a specific color that was fixed in each of the four quadrants. The borders were about an eighth of an inch, making them visually salient to participants. In other words, the top-left image always appeared with a red border, the bottom-left image had a green border, the top-right image had a yellow/orange border and the bottom-right had a blue border. During the first four trials, children saw two astronauts and two planets that were one of two colors, blue and orange.

Children heard simple phrases "*Find the [color] planet*" and verbally responded with the color border around the image. This continued until children identified each of the four objects on the screen. The verbal response was used to evaluate accuracy on the task.

After the practice trials, the experimental trials proceeded as described in Experiment 4.1. The critical difference between the earlier experiment and Experiment 4.2 is that during webcam eye-tracking, there was no fixation point to initiate the start of each new trial. Instead, after children provided a verbal response to identify the quadrant of the target image, they pressed the space bar to advance to the next trial.

Administration of webcam eye-tracking task. Children participated in this study either in person or remotely using webcam eyetracking. The administration procedures were similar to the description provided in Experiment 4.2 with the exception that we used webcam eye-tracking to collect gaze data for all participants. For remote visits, all participants were invited to an experiment session using a Zoom link and were guided through the experimental task by both a parent and research assistant on the virtual call. Further details for experimental setup and in person or virtual administration procedures can be found in Appendix J.

Data preparation and cleaning

Manual coding. Before webcam videos were independently coded for analysis, research assistants underwent an intensive training and calibration process. As described by Venker et al. (2020), the manual coding of eye-tracking experiments is by no means a trivial process. Hand coding requires extensive training, practice, calibration with a team of coders, and periodic review of the videos that have been coded. At the onset of training, research assistants were provided with a detailed coding guide that described the process of coding. This information included descriptions about codes that would be used, the coding software and its related functions, and visual examples of *typical* looking behaviors that would be observed while coding. A lead coder (i.e., a person that was experienced with hand coding and the coding software) was selected to provide research assistants with the scaffolding required to learn to become reliable coders.

The training procedures contained three stages: a) guided coding, in which trainees coded trials in the presence of a lead coder following an answer key, b) independent coding with an answer key accessible, and c) independent coding without an accessible answer key. This process was used to gradually fade the required level of support needed for reliable video coding; this helped coders smoothly transition from training to the calibration phase. Additional details about manual coding, training, and calibration procedures can be found in Appendix K.

Eye-tracking preparation and cleaning. Peyecoder was used to automatically extract all codes and convert information from frames to 33ms bins of time, an automatic adjustment provided by the program. Prior to extracting all of the data across participants, all test trials were independently examined by an independent coder and compared to "practice trials" to determine if recordings were mirrored. Using a built-in Peyecoder function, codes were transformed to accommodate mirrored recordings.

Data cleaning and screening followed the same procedure described in Experiment 3.1 (Erskine, 2022b). Children were excluded from all further analyses if they had fewer than 50% of reliable trials in a single block of eye-tracking. Data quality was then examined in region of time that began 250ms after the onset of the baseline region (*find the*) to 1500ms after the onset of the labeled referent (*socks*). This general window was selected to account for the time it takes children to program a saccade once they start to hear a sentence unfold.

Data cleaning procedures resulted in the retention of 244 out of 357 trials. Seven additional children were tested, but they had greater than 50% of missing data in at least one block of eye-tracking and were therefore excluded from all subsequent analyses. Among African American listeners, a more conventional criterion for missing data was used because the data were collected through webcam recordings which allowed for increased data retention (Erskine, 2022b; Venker et al. 2020). The proportion of missing data did not differ across Speaker Group Conditions, $F(2, 242) = .187$, $p = .65$. Table 6 and 7 provide a summary of the track loss analysis for participants before and after trial- or participant-level exclusion.

Results

Accuracy data was not obtained for all participants due to challenges related to remote testing or in person compliance. While procedures were created to record accuracy such as having colored borders around pictures, sometimes parents were not able to record children's performance on an alternative device or alternative attempts to acquire accuracy were disrupted by uncontrolled equipment challenges (i.e., children's responses were too soft to be record or the recording equipment failed to record a verbal response). For these reasons, we did not analyze or report response accuracy data. Instead, we complete gaze analyses across all trials that contained reliable data post screening procedures.

Table 6.
Summary of the track loss and sample analysis before trial-exclusion (n=31). AA is the abbreviation for African American listeners and EA is the abbreviation for European American listeners.

Speaker Group	Number of Trials	Mean Samples	Percent Track Loss (%)
Saw AA Speaker: Heard GAE	124	91	39.7
Saw AA Speaker: Heard AAE	123	90	41.8
Saw EA Speaker: Heard GAE	124	90	39.2

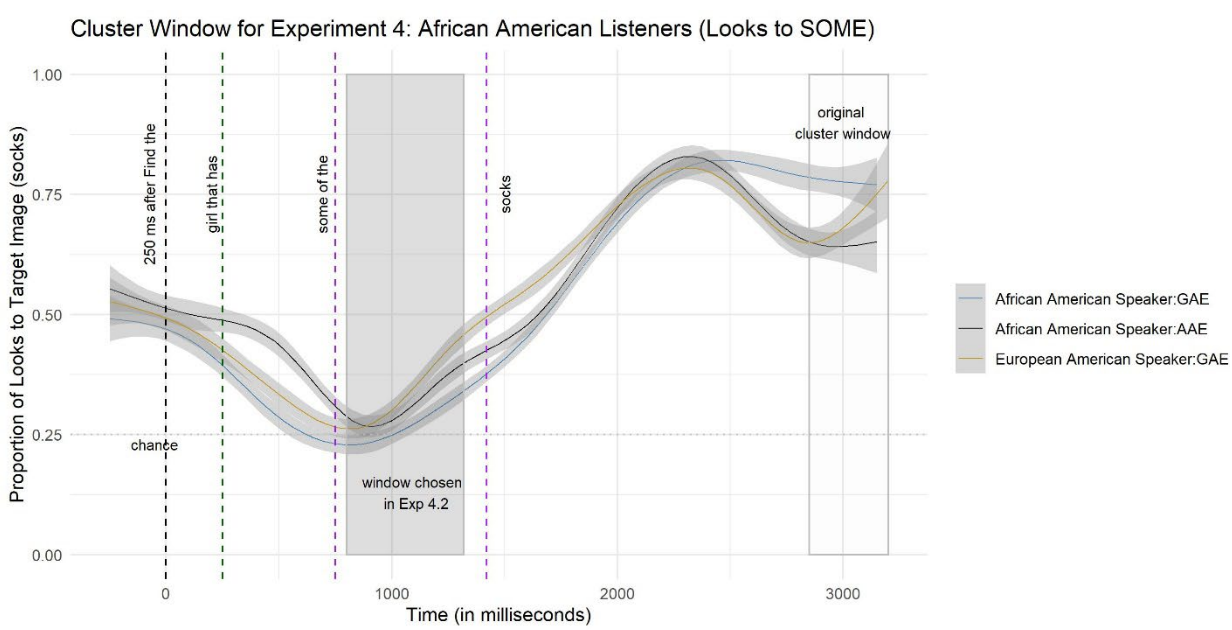
Table 7.
Summary of the track loss and sample analysis after trial-exclusion (n=25). AA is the abbreviation for African American listeners and EA is the abbreviation for European American listeners.

Speaker Group	Number of Trials	Mean Samples	Percent Track Loss (%)
Saw AA Speaker: Heard GAE	85	93	32.5
Saw AA Speaker: Heard AAE	80	93	29.8
Saw EA Speaker: Heard GAE	79	93	28.3

Cluster permutation analysis. Like Experiment 4.1, we used an empirical approach to determine the window of interest for evaluating effects of speak group on scalar inferencing. The cluster method identified four contiguous 50ms time bins that occurred long after the offset of target referent (socks), 3100ms-3450ms relative to the onset of the sentence at 0ms (i.e., the baseline region). This was not surprising because differences across conditions were largest at the end of the trial, a region of time which captured children's verbalized responses to the task (i.e., when children provided the label of the border color around the image of their choosing). However, from a conceptual point of view, it is unclear how to interpret looking patterns in this window of time. Language processing that is observed approximately 1500 ms after the onset of the noun (socks) most likely represents additional post-processing of auditory and visual information after scalar implicatures and verbal responses have been generated. From an analytical and practical point of view, analyzing looks in this window also makes it challenging to compare children in this study to previous studies on scalar implicatures (i.e., Huang & Snedeker, 2018; Grodner et al. 2010; and results in *Experiment 4.1*).

Since we are most interested in understanding how auditory and visual aspects of the speaker guide scalar implicatures, we used an earlier window of analysis, specifically the region of time that was identified in the cluster analysis in Experiment 4.1. As described earlier, this region began 50ms after the onset of the scalar quantifier (some) and ended 100ms prior to the target referent (socks).

Figure 5. Looks to target comparing the cluster region *selected* to the *original* cluster region identified by the cluster analysis. The grey window corresponds to the selected region, and the unshaded rectangle corresponds to the original cluster region identified by the cluster analysis.



Generalized linear mixed effects analysis and results. A generalized linear mixed effects model was fit to the cluster region to quantify the influence of Speaker Group and additional child-related measures of language and linguistic maturity on children's fixation to the correct interpretation (i.e., the image corresponding to the subset of referents). The dependent variable captures the log odds of looking at the target image. The predictor variables included fixed effects for Speaker Group (a contrast coded categorical variable), Vocabulary (as a centered continuous raw score on the PVT), Dialect Density (as a centered continuous measure from the DELV-ST) and Age (as a centered continuous variable). We include dialect density because African American children in this sample produced a broader range of NMAE features on the DELV-ST. Additionally, the model included two-way interactions between Speaker Group and all other aforementioned independent variables.

Contrast coding was used to evaluate the prediction that African American listeners would prioritize visual aspects of the speaker's identity (i.e., cues of racial background) above and beyond cues related to the speaker's language use. Recall, these predictions converge with results reported in Erskine et al. 2022b (Experiment 3.2) and hypotheses about children's lived experiences with individuals across different speaker communities. The reference level in each contrast is set to the speaker group condition in which listeners heard GAE paired with an African American photograph. The model evaluates two specific contrasts of Speaker Group. The first contrast corresponds to differences in children's fixation to the target image when hearing speakers of GAE that differed in race, either African American or European American. The second contrast corresponds to speakers of the same race that differ in the dialect spoken, either AAE or GAE. The intercept in this model represents the log odds of looking at the target image averaged over the three speaker group conditions for individuals of average age and vocabulary size. The R syntax is provided in Table 8.

We observed a significant interaction between Dialect Density and the second contrast of Speaker Group, indicating that as dialect density decreased, looks to the target for the second contrast (European American face/AAE dialect) increased, resulting in a larger difference between the first (African American face/AAE dialect) and second contrast (European American face/GAE dialect) for children with low dialect density. By contrast, children with high dialect density showed similar eye gaze patterns across the three Speaker Group conditions. We also observed a significant interaction between Age and the second contrast with Speaker Group indicating that as age decreased, looks to the target for the second contrast (European American face/AAE dialect) increased, resulting in a larger difference between the first (African American face/AAE dialect) and second contrast (European American face/AAE dialect) for children with low dialect density. Similarly, older children showed similar eye gaze patterns across the three Speaker Group conditions. All other main effects and interactions were not significant.

Table 8

Summary of parameter estimates for the mixed-effects logistic regression model fit to the gaze. Model was fit using maximium likelihood.

R Syntax: glmer (cbind(LooksTarg, LooksNotTarg) ~ Vocabulary (*centered raw score*) + Age (*centered*)+ Speaker Group (*contrast coded*) + Speaker Group*Vocabulary + Speaker Group*Age + (Speaker Group | participant) + (1 | trial)

Fixed Effects

	b	**SE**	**z**
Constant	-2.09	.68	-3.04**
Vocabulary (*NIH Toolbox PVT Raw Score*)	-.09	.58	-.16
Age (*In Months*)	.14	.56	.26
Dialect Density (*DELV-ST*)	-.13	.50	-.27
Speaker Group (Contrast 1)	.30	1.07	.28
Speaker Group (Contrast 2)	-.83	1.27	-.66
Vocabulary*Speaker Group (Contrast 1)	1.04	1.22	.85
Vocabulary*Speaker Group (Contrast 2)	-.94	.99	-.96
Dialect Density*Speaker Group (Contrast 1)	1.24	1.06	1.16
Dialect Density*Speaker Group (Contrast 2)	1.84	.88	2.08*
Age* Speaker Group (Contrast 1)	-.23	1.18	-.19
Age * Speaker Group (Contrast 2)	2.47	.96	2.57*

Random Effects

Parameter	**Variance**	**SD**	**Correlations**
Participant-level			
Constant	4.85	2.20	
Speaker Group (Contrast 1)	21.60	4.65	.50
Speaker Group (Contrast 2)	12.81	3.71	.57 .47
Trial-level			
Constant	4.34	2.08	

p <.001***, p < .01 **, p < .05*, $p \geq .05$ +

Figure 6. Looks to the target image across Speaker Group Conditions.

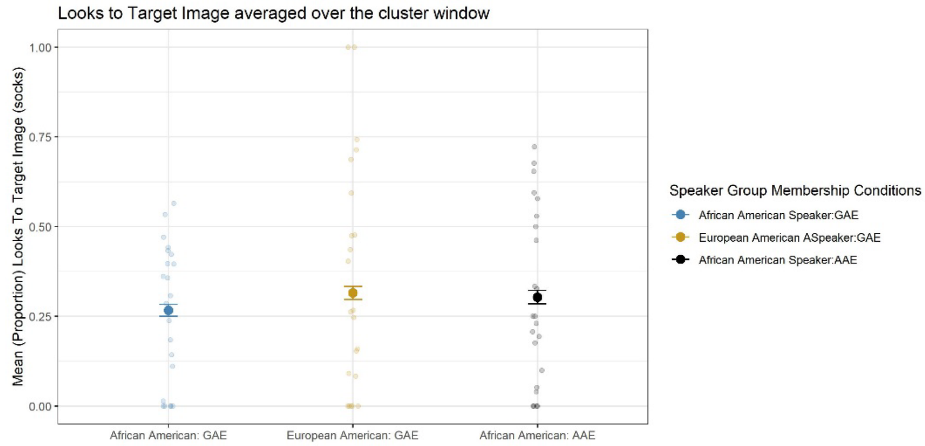


Figure 7. Looks to the target as a function of dialect density (median split). Across time (top plot) and averaged within the cluster region of analysis (bottom plot).

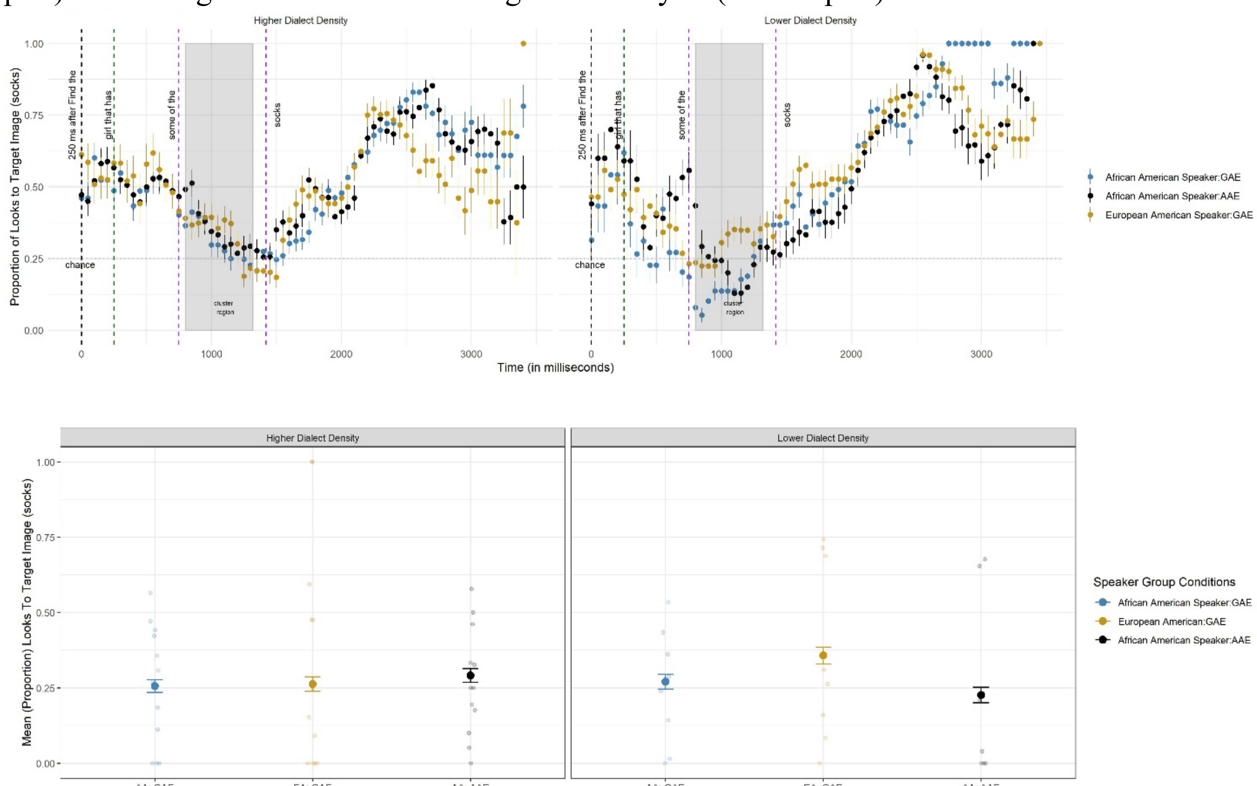
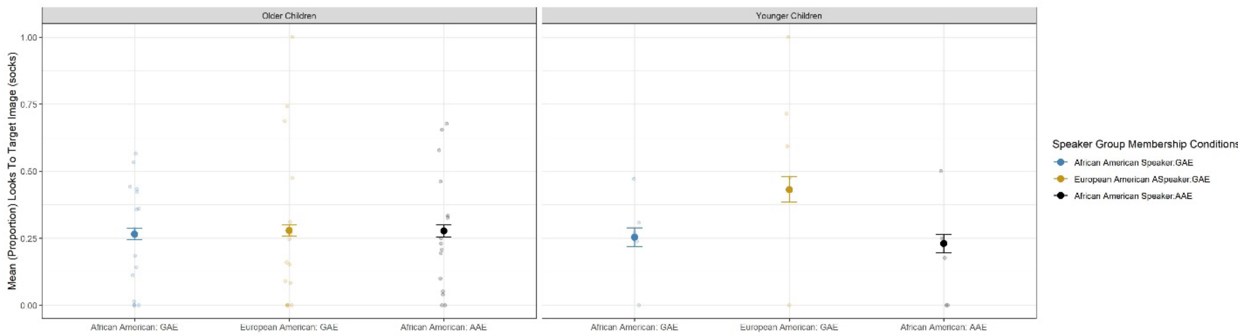


Figure 8. Looks to the target as a function of age (median split)



Discussion

Experiment 4.2 evaluates the effect of speaker identity on children's comprehension of the scalar implicature (*some, but not all*) among a group of listeners who have more variable experiences with speakers of dominant and/or stigmatized varieties of English (AAE and GAE). Overall, this study found few significant differences in children's ability to generate scalar inferences as a function of the experimental condition and the additional measures of language ability and cognitive maturity. Additionally, the results that were observed substantially differed from the initial predictions. The central finding of this study is that effects of speaker identity on spoken language comprehension tasks that require children to generate implicit or pragmatic interpretations are complex.

In the quantifier region (i.e., the region in which the scalar term *some* was heard), two patterns that varied as a function of children's nonmainstream dialect use emerged. Listeners with higher dialect density scores (i.e., AAE speakers) demonstrated fewer differences in their ability to generate implicit interpretations across the three speaker group conditions. By contrast, listeners with lower dialect density (i.e., primarily GAE speakers) showed an overall processing advantage for generating a scalar inference in the condition least expected (i.e., European American face/GAE dialect). Recall, in this study it was predicted that African American children would prioritize visual aspects of the speaker (i.e., racial cues) over language use (i.e., dialect-related cues) during comprehension tasks in which listeners were required to generate pragmatic interpretations for the scalar term *some.* Instead, this study found that that while all children eventually looked to the target image corresponding to the subset interpretation (i.e., generated a scalar inference) (see Figure 6), in the window of analysis (i.e., the quantifier region), speaker identity had more pronounced effects for speakers of low dialect density as compared to speakers with higher dialect density. That is, AAE-speaking children were close to chance in fixating to the target image across all speakers group conditions (regardless of race or dialect), while listeners who spoke primarily GAE showed a greater number of fixations to the target image when hearing GAE paired with a speaker of a different race (i.e., saw a European American face). To summarize, an effect of race was observed but only for children who were primary speakers of GAE and the effect was observed in the opposite direction than was anticipated.

The results of this experiment are puzzling and many factors may have contributed to the differences that were observed. First, the observed results may be capturing the range of complex experiences involving children's use of language, their ability to reason about speakers of similar and different backgrounds, and their overall experiences with how cues related to language and race interact across a variety of formal and informal communicative contexts. For example, it is plausible that the experimental measure provided an index of children's knowledge of how different types of language is conveyed across communities and in particular how conversational implicatures are used across speaker communities. Previous research has shown that specific types of pragmatic language use such as indirect requests are more likely to be observed in formal school contexts and among caregivers from European American families. Additionally, direct requests and prohibitions have been commonly observed among families from either lower SES backgrounds or African American families (Hoff-Ginsberg, 1994; Rowe et al., 2004). AAE-speaking listeners may have more communicative experiences in which direct or explicit meanings are commonly conveyed among other African American speakers regardless of dialect. Moreover, they likely have fewer interlocutory experiences with European American speakers of GAE in which there is more frequent use of implicit language such as indirect requests. Therefore, fewer looks to the target image across all three speaker conditions may be related to attenuated experiences in which listeners are able to quickly deduce that speakers across the three conditions intended to convey an implicit meaning (*some but not all*) in this task. This explanation which posits an interaction between experience with different speaker communities and pragmatic language use may also explain performance among children with lower dialect density. Recall, children with lower dialect density produced a greater number of fixations to the target image when hearing GAE paired with a photograph of a European American. This would suggest that it was easier for listeners to access and retrieve implicit interpretations in a dialect that might be familiar to children but from speakers who are different in race. This could be because GAE-speaking African American children have more experiences in which they hear European American speakers use implicit language to convey information (i.e., indirect requests) as compared to African American speakers who may use more direct language. This explanation is merely speculation because it is not entirely clear what is driving performance across this sample of African American listeners. However, these results do highlight the importance of understanding the contribution of speaker identity to pragmatic inferencing across a diverse group of listeners in the world.

Alternative explanations for these findings may be related to either the inclusion of vignettes in this study or noisy data due to the use of webcam eye-tracking. Before hearing the critical audio (*find the girl that has some of the socks*), children heard vignettes describing the images on the screen. Perhaps these vignettes provided listeners with an opportunity to obtain additional exposure to the different speakers, minimizing differences across speaker group conditions. This explanation may account for performance differences across AAE-speaking children, but it does not adequately explain differences among GAE-speaking African American children. Moreover, it seems peculiar that even a short duration of increased experienced listening to different speakers would result in GAE-speaking children receiving greater facilitation or benefit in the condition where listeners saw a European American photograph and heard GAE. Lastly, it is plausible that speaker identity effects are difficult to observe in this experiment which uses webcam eye-tracking. A critique which was raised also by Erskine (2022b) is that while web-cam eye-tracking assists with the recruitment of children from under-represented backgrounds, the data collected in remote contexts is noisy and therefore it is necessary to test relatively larger sample sizes and/or increase the number of trials within a single experiment (Anwyl-Irvine et al., 2020; Venker et al., 2020; Yang & Krajbich, 2020). Perhaps with a greater number of trials and participants, the results of this task might reveal clearer results regarding how speaker identity influences children's ability to generate scalar inferences in real time.

General Discussion

To summarize, in two experiments we provide preliminary evidence that European American and African American listeners demonstrate differences in their recruitment of speaker identity cues for inferring pragmatic interpretations. While European American listeners demonstrated an overall processing advantage for individuals that were most familiar or similar to them (i.e., speakers of the same race and dialect), African American listeners showed either no clear difference or showed the opposite pattern (i.e., increased fixations to the target image when hearing GAE paired with a photograph of a European American Speaker).

Together, these studies demonstrate that real-time integration of speaker identity cues during pragmatic inferencing is complex. While familiar speaker contexts can be facilitating for some children, this may not be a foregone conclusion for all children across different listener communities. Intriguingly, this is the first study to show that African American listeners perform either comparably or better at processing GAE relative to AAE. This result contrasts with previous studies which typically show an overall disadvantage for less familiar speaker or dialect contexts (i.e., GAE) (Charity et al., 2004; Edwards et al., 2014; NP Terry et al., 2016; Washington et al., 2018).

Nevertheless, these experiments do call into question the overall role of speaker identity in pragmatic language processing. Based on the pragmatic inferencing literature, we predicted that speaker identity would be highly salient and relevant for retrieving implicit interpretations. Additionally, we presumed that this would be particularly important for linguistic phenomena where there are strong requisites for the integration of the speaker's identity (i.e., scalar implicatures). In Experiment 4.1, the results seem to corroborate this assumption. Among European American listeners we saw evidence that familiarity positively influenced comprehension. Moreover, children did not show a dialect-based preference for facilitating comprehension. This finding neatly converges with speculations that reasoning about speakers and their intent is much easier when that person is from the same social group. Further, auditory and visual cues to speaker identity seemed to provide sufficient information for listeners to engage in this reasoning (i.e., listening to European American GAE speakers helped European American listeners generate scalar predictions with greater efficiency). However, Experiment 4.2 questions the validity of that assumption. Specifically, we see that familiarity did not successfully account for the performance observed among all African American listeners. This is the first among studies to assess the influence of socioindexical properties of the speaker on children's pragmatic reasoning of speakers of different races and dialects. For this reason, we modified conventional scalar implicature tasks to increase the likelihood that children would generate scalar inferences. Perhaps, these changes made the task too easy. If this is the case, then future studies may need to use more demanding pragmatic inferencing tasks to evaluate how properties of the speaker interact with the listener's ability to infer implicit as opposed to explicit meanings.

Taken together, Experiments 4.1 and 4.2, do not provide a clear answer to how speaker identity is used for pragmatic inferencing or deriving communicative intent in general. On the one hand, it is plausible that listeners' integration of linguistic and nonlinguistic speaker identity cues depends on the kinds of pragmatic inferences that are being made. In this paper, we focused on scalar quantifiers because it is a task that children ages 6-11 can reliably perform and does not necessarily require integration of broader world knowledge in ways that can drastically differ across communities of people. Perhaps, these socioindexical cues operate differently when the demands of the task are greater as in a more demanding scalar implicature task or when listeners need to interpret metaphors or derive other implicit meanings. On the other hand, it may be the case that trying to understand effects of speaker identity on spoken language comprehension is not as meaningful for evaluating relationships between dialect variation and spoken language comprehension. Though we observed differences in spoken language comprehension that are sometimes clear to interpretation, it is generally the case that all children regardless of their dialect status seem to comprehend the spoken language in these studies without significant challenges (see also Erskine, 2022b). It could be more important that we focus on contexts where children are learning new information as is suggested by more canonical perspectives embedded in epistemic trust frameworks (Shafto et al., 2012) or by Weatherhead et al.'s (2021) research on mutual exclusivity and word learning.

Previous studies (i.e., Erskine, 2022a, Erskine, 2022b) have been unable to isolate perceptual processing mechanisms from more socially-driven mechanisms that account for relationships between dialect variation and spoken language processing. Perhaps the limitations are rooted in the demands and objectives of previous experimental tasks. Pragmatic inferencing may provide a more direct assessment of communicative intent and how listeners engage with properties of the speaker to derive this intent. This paper makes two important contributions to the literature. First, it builds on previous literature that quantifies effects of speaker identity on spoken language processing by looking at a different aspect of language comprehension (i.e., listeners' comprehension of implicit meanings). Second, this study helps to disentangle the role of perceptual processing (which emphasizes the relevance of the auditory signal) from more socially-informed hypotheses about how listeners process language from different speakers (i.e., epistemic trust frameworks). Across two studies, we provide evidence that perceptual analysis on its own cannot adequately explain differences in spoken language processing, particularly for tasks that require the interpretation of implicit or pragmatically intended meanings.

Chapter 5: General Discussion

The primary objective of this dissertation was to consider the role of perceptual analysis hypothesis as a causal explanation for the relationship between dialect variation and spoken language comprehension. A growing segment of the dialect variation literature has asserted that linguistic differences between AAE and GAE create additional challenges in mapping an incoming auditory signal to listeners' mental representations of the signal (i.e., ***the perceptual analysis hypothesis***). However, very little research has been done to empirically evaluate the validity of this hypothesis as a causal link between dialect use and linguistic processing. Using the *visual world paradigm*, a measure that is sensitive to the real-time demands of spoken language processing (i.e., the relative ease of word recognition and/or the consideration of possible interpretations), four experiments were used to examine the contribution of dialect variation to spoken language comprehension. Further, these experiments included stimuli that varied in linguistic complexity and access to cues about the speaker's identity. The role of dialect variation was examined in three linguistic contexts: a) word recognition in simple phrases, b) word recognition in longer utterances that contained semantically facilitating or neutral verbs, and c) sentences with implicit, pragmatic interpretations. Additionally, this dissertation compared children's sensitivity to dialect variation in two types of visual world paradigm experiments: a) access to auditory cues only (Chapter 2 [Erskine, 2022a]) and b) access to both auditory and visual cues about the speaker's identity (Chapter 3 [Erskine, 2022b] and Chapter 4 [Erskine, 2002c]). Finally, in contrast to many prior studies, the participants in each study included a substantial percentage of African American children who were GAE speakers.

Across each of the four experiments, some of the findings supported a perceptual analysis explanation while other findings could not be easily explained through perceptual analysis lens. To start, chapter 2 (Erskine, 2022a) showed that for linguistic contexts where differences between dialects varied in perceptual salience (i.e., attenuated [word recognition in simple phrases such as *find the bee*] vs. highly salient differences between AAE and GAE [word recognition in longer utterances]), effects of dialect variation on spoken language comprehension were muted. That is, when auditory information alone was available to listeners, the effects of dialect variation were most robust among listeners with the least amount of familiarity with a dialect that was not their primary dialect (i.e., European American GAE listeners hearing AAE). Further, effects of dialect variation occurred only in longer sentences where facilitating semantic information was absent (i.e., sentences with neutral verbs such as *choose the book).* By contrast, African American children who were speakers of AAE did not show differences in recognizing words in simple or more demanding linguistic tasks regardless of the dialect that the stimuli were presented in (AAE or GAE). The results of Erskine (2022a), and in particular those for European American children who were GAE speakers, provide converging evidence that if listeners are able, they will rely on higher-level sources of information such as semantic context to facilitate comprehension of a dialect that differs from the listener's primary dialect, as was previously observed by Lev-Ari (2015). Hence, this paper provides additional evidence that school-age children (5- to 9-years-olds), like adults, leverage higher-level linguistic context to anticipate upcoming information which aids overall spoken language comprehension. It also shows that this ability can become hampered in linguistic environments in which listeners have less contextual information to rely on. Erskine (2022a) only examined semantic facilitation within a single sentence, but one can imagine that higher-level cues from the broader linguistic or non-linguistic environment such as the visual scene or knowledge-based cues about the speaker will also facilitate comprehension of different dialects.

This was directly tested in Chapter 3 (Erskine, 2022b) and 4 (Erskine, 2022c) which examined children's integration of linguistic and non-linguistic speaker identity cues in two spoken language contexts: literal/explicit semantic comprehension and pragmatic/implicit comprehension. In each study, it was observed that when auditory and visual speaker identity cues were available, the effects of dialect variation on spoken language comprehension depended on children's primary dialect, sociocultural experiences, and the overall objectives of the task. When the task required listeners to recognize familiar words in semantically facilitating sentences (i.e., predictive verbs), Erskine (2022b) found evidence that European American GAE-speaking children prioritized cues related to the speaker's dialect as opposed to race which resulted in rapid and efficient word recognition of other GAE speakers regardless of race. By contrast, African American listeners demonstrated a trend in which racial cues were more informative for efficient comprehension. Moreover, these effects were more pronounced among children with higher rates of AAE features (i.e., AAE-speaking African American children). At first glance, these results indicate that accent may be a more useful cue for reasoning about speakers for European American children but less useful for African American children who speak a stigmatized dialect of English. This interpretation would corroborate findings observed in the social psychology literature (Kinzler et al., 2009). However, Erskine (2022c) provide conflicting results. When listeners had to infer an implicit interpretation for the scalar quantifier *some*, European American children seemingly changed strategies by prioritizing cues to race over and above dialect and African American children demonstrated either no differences in comprehension among speakers or a slight advantage for speakers of a different race and dialect. The results of this final study showed that while European American listeners showed an advantage in processing scalar implicatures for individuals most similar to themselves, African

American listeners showed no such advantage and in fact children who were more GAE-aligning in their use of American English, showed an advantage for European American speakers of GAE.

The results across the two latter studies suggest that effect of speaker identity cues on spoken language processing are not straightforward. Overall, the evidence shows that children flexibly recruit knowledge about the speaker in different ways to guide their comprehension. For European American children, accent/dialect may be a salient cue for comprehending information that is immediately available through semantic analysis (i.e., recognition of highly familiar words) but less helpful for tasks that require pragmatic inferencing because these tasks place greater demands on the listener's ability to reason about the speaker and their communicative intent (i.e., implicit meaning that is conveyed through words). For African American children, the results may depend on factors that play a more prominent role in minoritized communities. For example, across the last two studies, African American children's performance was best characterized by their primary dialect use. African American children who were AAE-speakers (i.e., had high dialect density on the DELV-ST) prioritized cues to race during tasks that required semantic analysis only, but showed much less pronounced effects of speaker identity when they had to access an intended meaning. By contrast, African American children who were GAE-speakers (i.e., had low dialect density on the DELV-ST) showed the opposite trend in which no differences across speakers were observed during tasks that relied on semantic analysis, but showed a processing advantage for the condition in which GAE was paired with a European American face when they had to access an intended meaning. The different results among European American children, African American children who spoke GAE, and African American children who spoke AAE may best be explained by considering children's socialized experiences with how language is used across different speaker communities. This perspective assumes that effects of speaker identity are not as simple as crude divisions about dialect and/or race-related familiarity. Instead, spoken language processing is informed by how speaker identity cues are linked to different social or communicative contexts for different listener groups, which also includes experiences about language use across speakers.

While this dissertation has provided some answers regarding the validity of the perceptual analysis, there are a few limitations that should be addressed in future studies. One obvious limitation is the relatively small sample size for each study, a consequence of conducting research with human subjects during a global pandemic. A second important limitation of the current studies is that all studies included only a single speaker per condition, making generalization more difficult since there may have been talker-specific effects. (For example, the African American GAE speaker had an especially clear and resonant voice, as noted in Erskine, 2022b.) Including multiple speakers per condition is problematic, however, because it is not clear how the integration of speaker identity during the comprehension of different dialects will vary as a function of hearing a single talker vs. multiple talkers. That is, it will be challenging to disentangle more general multiple-speaker effects from dialect- or race-related effects. Previous studies have shown that including stimuli from multiple speakers as compared to a single speaker sometimes make spoken language comprehension more challenging and sometimes make it easier, depending on the task (word recognition and sentence processing), listening environment (quiet vs. noisy) and whether there is additional visual information about the speaker (Nygaard et al., 1994). For example, audio-visual speaker contexts with multiple talkers have led to poorer perceptual processing as compared to hearing and seeing a single talker (Heald & Nusbaum, 2014). This suggests that the role of speaker identity is complex and a variety of speaker contexts need to be evaluated to understand the benefits or costs associated with dialect variation in socially-enriched speaker contexts. Another limitation of this dissertation is that Chapter 4 (Erskine, 2022c) includes a single measure of pragmatic inferencing (i.e., scalar implicatures) as opposed to a variety of pragmatic language contexts. This is relevant because the results of Erskine (2022c) could be misconstrued to suggest that African American children either have infrequent linguistic experiences with implicit language use with other African Americans or that implicit language use is predominantly a feature of European American speakers of GAE. In contrast to this interpretation, Erskine (2022c) posits that the results observed among African American listeners were attributed to the specific pragmatic context examined (scalar quantifiers). Erskine (2022c) evaluated children's sensitivity to the speaker's communicative intent in a task that children could reliably perform and one that differs from tasks which only require semantic analysis because of the greater demands to integrate information about the speaker's identity degree (Huang et al., 2009). However, research that examines children's comprehension of scalar quantifiers may not best for investigating questions about how dialect variation influences the comprehension of pragmatic language. One reason for this is because researchers have limited understanding of pragmatic language use among minoritized communities such as African Americans or speakers of African American English. In a recent systematic synthesis of pragmatic language use among African American families (children and caregivers), Hyter et al. (2015) found that most of the research dating back to the 1950's focused on the following topics: a) discourse coherence, b) use of AAE in narrative production, c) use of conjunctions, and d) selection of creative writing topics among African American children (i.e., examining whether children wrote stories recounting lived experiences or created stories featuring morally-centered characters). Interestingly, none of the studies reported on children's use of implicit language to articulate ideas or on caregivers' use of pragmatic language in structured or unstructured language sampling tasks. This body of literature points to a large gap in our understanding about pragmatic language development in minoritized communities. Until there is additional research about pragmatic language development across underrepresented linguistic communities, research on dialect variation and the semantic-pragmatic interface should focus on aspects of pragmatic language use that are less culturally specific and perhaps, the most relevant for academic achievement (i.e., comprehension of metaphors or other kinds of indirect uses of language).

Despite some of these limitations, the central takeaway from all four experiments is that perceptual analysis provides a very limited account for the influence of dialect variation on spoken language comprehension. This conclusion is partly supported by the restricted contexts in which effects of dialect variation were observed. When dialect differences were heard in socially-attenuated contexts (i.e., dialect variation heard in the absence of other social cues about the speaker [Chapter 2]), listeners leveraged both their linguistic knowledge and experiences to process language. These findings are consistent with the perceptual analysis hypothesis which takes into account both the linguistic context (e.g., semantically facilitating information or conversational topics) and listeners' experiences (e.g., familiarity with dialects heard). However, these findings also suggest that difficulties with linguistic processing may not completely hamper listeners' comprehension of dialect variation, at least not to the degree that sociolinguists and applied scientists have implied (JM Terry et al., 2010; Edwards et al., 2014; Washington et al., 2018).

Relatedly, effects of dialect variation across some of the studies were consistently observed relatively late in processing. Recall, Chapter 2 (Erskine, 2022b) and Chapter 3

(Erskine, 2022c) analyzed gaze patterns in windows that occurred approximately 500 to 1350 ms after the onset of the verb. Moreover, peak looks to the target image occurred 4000-4500ms after the onset of the sentence. Delays in processing may have been partly attributed to including longer sentences (as has been shown by Blomquist et al., 2021). However, the delays observed in this dissertation may also reflect later integration of sociolinguistic information; these effects are usually late arriving relative to phonological and linguistic factors driving real-time perceptual analysis of the speech signal. This later integration is not easily captured by a perceptual processing account which would predict time course differences that are much earlier in the time course or at least more closely linked to the regions of lexical ambiguity.

Another reason to be wary of the perceptual analysis hypothesis is that it does not account for the influence of non-linguistic properties of the speaker. First, in contrast to Chapter 2 (Erskine, 2022a), Chapter 3 (Erskine, 2022b) posits that even the support provided by facilitating semantic information can be dampened in the presence of non-linguistic information about the speaker's identity. Recall, while no effects of dialect variation were found in predictive verb contexts (*kick the ball*) in Chapter 2, more prominent effects were found in Chapter 3 (Erskine,2022b) which paired a face with the auditory stimuli (i.e., European American listeners showed poorer comprehension of African American faces paired with AAE and African American listeners with higher dialect density showed poorer comprehension for GAE when it was paired with a European American face). Additionally, Chapter 4 (Erskine 2022c) presents further evidence that dialect is not the only signal listeners use to guide spoken language comprehension, particularly in cases that require pragmatic interpretations. Both European American and African American GAE-speaking children prioritized race-related cues over dialect when generating scalar implicatures. Moreover, effects of speaker identity occurred in the expected direction for GAE-speaking European American children but in the opposite direction for GAE-speaking African American children (i.e., both groups showed more efficient scalar inferencing for European American faces paired with GAE).

While the perceptual analysis hypothesis makes predictions about the salience of the signal in different linguistic contexts, it does not make strong predictions about the role of speaker identity during spoken language processing. Moreover, if effects of dialect variation were driven primarily by linguistic differences in the auditory signal, then non-linguistic speaker identity cues should not have influenced children's word recognition. The results of the last two studies provide evidence that conflicts with such predictions. Therefore, it cannot adequately explain the results that were found in previous studies on speaker identity and spoken language processing or the results found in Chapter 3 (Erskine, 2022b) or Chapter 4 (Erskine, 2022c). Causal explanations for the relationship between dialect variation and spoken language comprehension must be able to explain *how* and *when* spoken language processing will be influenced by dialect variation in a variety of communicative contexts including: a) different sociolinguistic environments (i.e., the absence or presence of nonlinguistic information about the speaker's identity) and b) different linguistic contexts (i.e., word recognition in literal semantic or semantic-pragmatic inferencing contexts). A clear pattern across the four studies is that processing dialect variation in real time cannot be simply characterized as challenges related to perceptual analysis. The fact that perceptual analysis difficulties arise under very limited circumstances (or at least the restricted contexts evaluated in this dissertation) raises concerns about the validity of the perceptual analysis hypothesis as an adequate account for challenges that African American children experience in academic contexts. Encoding linguistic signals are important for linguistic processing, but this is not the only challenge listeners face when mapping variable speech signals to stored mental representations. Across many communicative contexts, listeners encode linguistic information that co-occur with additional nonlinguistic cues about the speaker. In Chapter 4, Erskine (2022c) shows that the integration of auditory and visual cues related to the speaker's identity can have meaningful influence on children's spoken language comprehension. Further, this influence is strongly correlated with listeners' lived experiences.

To date, an account that posits direct relationships among speaker identity, dialect variation, and spoken language comprehension has not been advanced. However, this dissertation in combination with previous research (McGowan, 2015; Munson, 2011; Staum Casasanto, 2008; Weatherhead & White, 2018; Weatherhead et al., 2021) suggests that there is a strong need for causal explanations that consider how different communicative contexts, and in particular how contexts that go beyond the linguistic signal, support dialect variation processing. Again, this is an inherent limitation of the perceptual analysis hypothesis more generally. One of the central arguments of Erskine (2022b) and Erskine (2022c) is that processing linguistic differences between AAE and GAE inherently engages the listeners' reasoning about linguistic and nonlinguistic aspects of the speaker's identity. That is, comprehension of dialect variation is a highly social behavior that results from complex interactions between the speaker's linguistic choices and the listeners' integration of the information provided both by the linguistic and non-linguistic properties of the speaker (i.e., speaker identity). Furthermore, dialects like AAE are usually nested in communities of talkers that are African American. Finally, the communicative environments in which speakers of AAE and GAE exchange information usually occur in contexts where listeners can avail themselves to both the auditory and visual properties of the speaker. Yet, a dearth of evidence exists in which research combines this social awareness of dialect use with experiments designed to investigate children's comprehension of dialect differences.

This gap in our understanding highlights an opportunity to conduct research that is related to the unanswered question of *why* speaker identity is relevant for spoken language comprehension. Much of the existing work on speaker identity has focused on listeners' sensitivity to these cues, but few studies have articulated a causal explanation that reasonably connects the influence of speaker identity to processes that underlie listeners' derivation of meaning from words, explicit or implicit. One candidate framework that was briefly highlighted in Erskine (2022b) as an explanation for the importance of speaker identity for spoken language processing is the epistemic trust hypothesis. Researchers in computer science, education, and word learning have proposed the epistemic trust hypothesis as a way to understand why aspects of speaker's identity (i.e., epistemic cues) are relevant for learning (Shafto et al., 2012). Epistemic trust provides a theoretical framework to explain the task children face in learning from different individuals about the world around them. From the perspective of the learner, it is necessary to differentiate individuals who provide reliable or trustworthy information from individuals who are less reliable or provide less relevant information (Mascaro & Sperber, 2009; Pasquini et al., 2007; Corriveau, Meints & Harris, 2009; Corriveau, Fusaro & Harris, 2009; Corriveau & Harris, 2009a, 2009b; Clement et al., 2004; Harris & Corriveau, 2011; Jaswal et al., 2010; Jaswal & Neely, 2006). To identify potential informants, the epistemic trust hypothesis posits that learners must identify relevant epistemic cues that determine whether a person is likely to be a reliable informant or not. Learners are most likely to consider a speaker as a reliable informant when the speaker comes from the same community as them, as determined by epistemic cues such as race/ethnicity and accent/dialect.

Studies from this literature have shown that children tend to show a preference for "familiarity" or “lived experiences” during learning. That is, previous studies have shown that children prefer to learn new words from caregivers as opposed to strangers (Barry-Anwar et al., 2017) and they will choose to be friends with or learn novel toy functions from speakers with familiar or native accents as compared to less familiar/non-native accents (Kinzler et al., 2009, Kinzler et al., 2011). These results are consistent with reports of a familiarity preference among children in the spoken language comprehension literature. For example, prior studies have shown that words are easier to *recognize* when they are produced by caregivers as compared to strangers or by individuals with familiar/native accents as compared to non-native accents. While the epistemic trust framework has primarily focused on the challenges associated with learning, it is plausible that epistemic cues related to the speaker’s identity guide spoken language comprehension -- just as prior experiences with different speakers (and the cues that inform assumptions about speakers) guide learning and other social decisions or preferences. However, deeper consideration of this framework for spoken language comprehension requires researchers to establish stronger working hypotheses about the ways in which epistemic cues assist listeners’ reasoning about the speaker. This is especially important because spoken language comprehension, which often involves the recognition of known words, may fundamentally differ from processes that facilitate word learning.

This leads to several important questions for future research. For example, is it the case that epistemic trust during spoken language comprehension helps to guide the listener’s reasoning about the reliability or credibility of the speaker as has been purported by research in word learning? And does the likelihood of informant credibility facilitate fast and accurate comprehension as it does word learning? Additionally, is informant reliability more important for less familiar words (e.g., low frequency words) as compared to highly familiar words? Alternatively, is it possible that epistemic trust in the context of spoken language comprehension is more related to the listener's ability to reason about the reliability of the signal itself as opposed to credibility of the speaker? Or is it the case the reliability of the signal and speaker interact such that if there is a mispronunciation or noise in the signal, can the listener "fill in" the missing information more easily if the informant is reliable? These questions capture the importance of understanding the relationships between epistemic trust and spoken language comprehension. Further, they highlight a much needed step which is to understand how to operationalize epistemic trust and define the parameters of trust that are most relevant for processing speech signals. Once some of these gaps in our understanding have been addressed, epistemic trust, which accounts for the influence linguistic and nonlinguistic properties of the speaker, may be a useful framework for investigating the effects of dialect variation in academic contexts.

Finally, this research provides at least one practical suggestion that can be implemented in applied education settings. One of the critical takeaways from this dissertation is the robust influence of *familiarity* and *experience* on children's comprehension of different dialects when paired with or without visual information about the speaker's face. That is, all three experiments showed that children were least efficient in recognizing words if they had fewer experiences with a speaker's race and/or dialect. Moreover, poorer word recognition in the aforementioned social and linguistic contexts was more likely to be observed among children who attended schools with more homogenous demographics (i.e., schools that were primarily attended by either African American or European American students). This suggests that social environments in which children have increased exposure to and experience with different cultural and linguistic practices are critical for mitigating potential difficulties in understanding speakers of less familiar dialects. This suggestion is also corroborated by a growing number of studies that have shown that exposure to language variation through explicit or implicit training greatly improves children's and adults' linguistic processing of unfamiliar accents or dialects (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Kutlu et al. 2022).

One way to increase children's familiarity with less familiar dialects is to provide them with greater opportunities for bi- or multi-dialectal experiences in the classroom. This can be accomplished by including texts in the language arts curriculum that are written by authors of different cultural and linguistic backgrounds. In fact, several educators have encouraged academic institutions to modify curricula to include more diverse literature in classroom contexts and in school libraries (Kibler & Chapman, 2019; Sharma & Christ, 2017; Shrodt et al., 2015). These texts have been used to introduce children to differences in prose and narrative or grammatical structures that can vary across languages and dialects. Integrating more diverse literature to children may also be crucial for capturing and exposing children to meaningful and naturalistic forms of language variation that exist within our society.

In addition to expanding children's literary experiences, this research provides futher support for the importance of having greater representation of educators from a variety of cultural or linguistic backgrounds. Sociolinguists have historically documented how language practices and use can vary substantially across regions and cultures (Labov, 2011, 2012; Mallinson, 2015). Representation among educators also provides organic ways to capture sources of language and cultural variation in classroom contexts. Moreover, this is a necessary step for providing children with greater exposure and meaningful language experiences with speakers of different races and language backgrounds.

However, we must also consider the extent to which the suggestion proposed above will equally benefit children regardless of the primary dialect spoken. That is, increasing the diversity of texts in the classroom may be the most helpful for children who primarily speak GAE (or other dominant dialects), but have more limited benefits for speakers of AAE (or other nonmainstream/stigmatized dialects). This is because many classrooms in the United States focus on providing instruction and language/literacy experiences that are centered around GAE. In fact, children's use of nonmainstream dialects are often either discouraged or relegated to creative writing. Therefore, the majority of children's academic experiences are in GAE (the dialect that is acquired in school for many speakers of AAE). In this case, for children who speak stigmatized or marginalized dialects of English, solutions based on increasing familiarity and experience with GAE may be less advantageous. In such instances, it may be more important to consider alternative approaches.

For example, it may be vital to consider connections among social identity, social relevance, and language use. It is plausible that children's experience with variation across dialects needs to be rooted in more meaningful social interactions. This is because the use of GAE may have specific social entailments that influence children's adaptation. For example, for many speakers of AAE, productive use of GAE occurs in more formal contexts such as in school or specific work-related environments. In these environments there is a greater expectation and demand to use GAE because of either institutional norms or the general distribution of speakers that are likely to exist in those specific social contexts (i.e., mostly speakers of GAE). For some children, these more formal contexts may have less social relevance or social capital which might impact children's development of flexible or tolerate representations of GAE. This explanation is largely speculative, but presents an idea that is open to further investigation.

Moreover, it is an explanation that might be more closely linked to the importance of epistemic trust and how this framework operates in communicative contexts more broadly.

Appendix A: Supplementary model outputs and additional summary statistics from Experiment 1 (Spoken Word Recognition) and Experiment 2 (Semantic Prediction)

Table A1. Summary of parameter estimates for the mixed-effects logistic regression model fit to the gaze data from **Experiment 1** (Spoken Word Recognition). These models were fit using maximium likelihood. Vocabulary is measured using the Expressive Vocabulary Test-Second Edition (Standard Score).

R Syntax: cbind(LooksTarg, LooksNotTarg) ~ Dialect Heard*Linguistic Group + Vocabulary*Maternal Education Groups + (Dialect Heard | Participant)

Fixed Effects			
	b	**SE**	**z**
Constant	.28	.10	2.77 **
Dialect Heard	-.08	.08	-.99
Linguistic Group	-.17	.16	-1.09
Vocabulary (EVT-2 SS)	.34	.07	4.43 ***
Maternal Ed (Mid)	.13	.16	.81
Maternal Ed (High)	.15	.16	.95
Dialect Heard*Linguistic Group	.08	.10	.77
Vocabulary * Maternal Education (Mid)	-.16	.17	-.91
Vocabulary * Maternal Education (High)	-.25	.12	-2.17 *
Random Effects			
Parameter	**Variance**	**SD**	**Correlations**
Constant	.11	.32	
Dialect Heard	.12	.35	-.50

p <.001***, p < .01 **, p < .05* , p ≥ .05 +

Table A2. Correlations across child-level variables in Experiment 1

	1	2	3	4	5	6	7	8
1. Eye-tracking Mean Prop Looks to Target Image	-							
2. Dialect Condition	-.03	-						
3. Listener Group	.15 ***	-.01	-					
4. Vocabulary (EVT-2 GSV)	.23 ***	.00	.68 ***	-				
5. Vocabulary (PPVT-4 GSV)	.22 ***	-.01	.70 ***	.91 ***	-			
6. Listener's Age in Months	.15 ***	.00	.16 ***	.41 ***	.46 ***	-		
7. Female	-.04 +	.01	-.18 ***	-.21 ***	-.15 ***	.03	-	
8. Maternal Education Level	.16	-.01	.81 ***	.68 ***	.67 ***	.18 ***	-.16 ***	-

*** $p < .0001$, + $p < .08$

Table A3. Summary of parameter estimates for the mixed-effects logistic regression model fit to the gaze data excluding age from **Experiment 2**: Semantic Prediction. These models were fit using maximium likelihood. Vocabulary is measured using the Peabody Picture Vocabulary Test-4th Edition using the Standard Scores (SS) which are age normalized.

R Syntax: cbind(LooksTarg, LooksNotTarg) ~ Semantic Condition + Dialect Density + Age + Dialect Heard + Race + Vocabulary (SS) + Semantic Condition*Dialect Heard + Semantic Condition*Vocabulary + Dialect Heard*Vocabulary + Semantic Condition*Dialect Heard *Vocabulary + Dialect Density*Dialect Heard + Semantic Condition*Dialect Density *Dialect Heard + (Semantic Condition | Item) + (Dialect Heard + Semantic Condition| Participant)

Fixed Effects

	b	**SE**	**z**
Constant	.55	.13	4.44 ***
Semantic Condition	-.92	.00	-162.79***
Dialect Density (*DELV-ST*)	-.12	.15	-.79
Dialect Heard	-.02	.00	-5.01***
Race	.05	.22	.23
Vocabulary (*PPVT-4 GSV*)	.12	.11	1.54
Semantic Condition * Dialect Heard	-.18	.01	-16.81***
Semantic Condition * Dialect Density	.08	.00	8.44***
Semantic Condition * Vocabulary	.00	.00	.47
Dialect Heard * Vocabulary	.09	.00	11.99***
Dialect Heard * Dialect Density	.11	.00	11.55***
Semantic Condition *Dialect Heard * Vocabulary	-.11	.01	-7.40***
Semantic Condition *Dialect Density * Dialect Heard	.10	.01	5.33***

p <.001***, p < .01 **, p < .05* , p ≥ .05 +

Random Effects

Parameter	**Variance**	**SD**
Participant-level Intercept	.16	.40
Item-level Intercept	.11	.33

Table A4.
Summary of parameter estimates for the mixed-effects logistic regression model fit to the gaze data excluding the NMAE Listeners (n=23 GAE Listeners ONLY). These models were fit using maximium likelihood. Vocabulary is measured using the Peabody Picture Vocabulary Test-4th Edition using the Growth Scale Value scores.

R Syntax: cbind(LooksTarg, LooksNotTarg) ~ Semantic Condition + Dialect Density + Age + Dialect Heard + Race + Vocabulary (GSV) + Semantic Condition*Dialect Heard + Semantic Condition*Vocabulary + Dialect Heard*Vocabulary + Semantic Condition*Dialect Heard *Vocabulary + Dialect Density*Dialect Heard + Semantic Condition*Dialect Density *Dialect Heard + (Semantic Condition | Item) + (Dialect Heard + Semantic Condition| Participant)

Fixed Effects

	b	**SE**	**z**
Constant	.72	.17	4.09 ***
Semantic Condition	-1.18	.20	-5.72***
Dialect Density (*DELV-ST*)	-.001	.29	-.004
Age (*In Months*)	.35	.10	3.52***
Dialect Heard	-.04	.14	-.30
Race	-.02	.26	-0.06
Vocabulary (*PPVT-4 SS*)	.07	.16	.47
Semantic Condition * Dialect Heard	-.13	.02	-7.08***
Semantic Condition * Dialect Density	-.47	.29	-1.60
Semantic Condition * Vocabulary	-.38	.16	-2.32*
Dialect Heard * Vocabulary	.11	.14	.75
Dialect Heard * Dialect Density	.08	.25	.30
Semantic Condition *Dialect Heard * Vocabulary	-.03	.02	-1.22
Semantic Condition *Dialect Density * Dialect Heard	.29	.04	7.44***

p <.001***, p < .01 **, p < .05* , p ≥ .05 +

Random Effects

Parameter	**Variance**	**SD**	**Correlations**	
Participant-level Intercept	.26	.51		
Participant slope: Dialect Heard	.17	.42	-.68	
Participant slope: Semantic Condition	.24	.49	-.20	-.04
Item-level Intercept	.21	.46		
Item slope: Semantic Condition	.22	.47	-.58	

Table A5.
Correlations across child-level variables in Experiment 2

	1	2	3	4	5	6	7	8
1. Vocabulary (PPVT-4 GSV)	-							
2. Semantic Condition	.00	-						
3. Dialect Heard	.01	.04***	-					
4. Dialect Density	-.58***	.02	.00	-				
5. Eye-tracking Mean Prop Looks to Target Image	.08***	-.20***	-.02*	-.09***	-			

6. Listener's Age in Months	.61***	.00	.01	-.25***	.08***	-		
7. Female	-.06***	.00	.00	-.05***	-.02***	.03***	-	
8. Race	.64***	-.01+	.01	-.74**	.09***	.32***	-.12***	-

*** $p < .0001$, * $p < .05$, + $p < .10$

Appendix B: Auditory Stimuli List Experiment 1 and Experiment 2

Table B1. Stimuli list in Experiment 1 by word group.

Target Referent	**Semantic Foil**	**Phonological Foil**	**Unrelated Foil**
Bear	Horse	Bell	Ring
Bee	Fly	Bear	Heart
Bread	Cheese	Bear	Vase
Cheese	Bread	Shirt	Van
Dress	Shirt	Drum	Swing
Drum	Bell	Dress	Sword
Heart	Ring	Horse	Bread
Horse	Bear	Heart	Pan
Kite	Flag	Gift	Shirt
Pan	Spoon	Pear	Vase
Pear	Cheese	Pen	Ring
Pen	Sword	Pear	Van
Ring	Dress	Swing	Flag
Shirt	Dress	Cheese	Fly
Spoon	Pan	Swan	Drum
Swan	Bee	Spoon	Bell
Swing	Kite	Spoon	Heart
Sword	Pen	Swan	Gift
Van	Horse	Pan	Sword
Vase	Gift	Van	Swan

Table B2. Stimuli list in Experiment 2

Dialect Heard Condition	Sentences containing **predictive** verbs	Sentences containing **neutral** verbs	AAE Features
AAE	Alyssa watchin' her broth(schwa) blow thehorn.	Alyssa watchin' her broth(schwa) choose the horn.	Zero-marked auxiliary (is) Derhotacized Vocalic R Alveolarization
GAE	Alyssa is watching her brother blow the horn.	Alyssa is watching her brother choose the horn.	
AAE	David and Tyler helpin' they sist(schwa) climb the tree.	David and Tyler helpin' they sist(schwa) fin' the tree.	Zero-marked auxiliary (are) Possessive (they)

GAE	David and Tyler are helping their sister climb the tree.	David and Tyler are helping their sister find the tree.	Derhotacized Vocalic R Zero-marked final consonant cluster Alveolarization
AAE GAE	They was watchin' they cousin drink the milk. They were watching their cousin drink the milk.	They was watchin' they cousin buy the milk. They were watching their cousin buy the milk.	Subject-verb-agreement (was/were) Possessive (they) Alveolarization
AAE GAE	Molly and Sue was watchin' Kate eat the sandwich. Molly and Sue were watchin' Kate eat the sandwich.	Molly and Sue was watchin' Kate fin' the sandwich. Molly and Sue were watching Kate eat the sandwich.	Subject-verb-agreement (was/were) Zero-marked final consonant cluster Alveolarization
AAE GAE	Tyrone and Sam watchin' they broth(schwa) fly the kite Tyrone and Sam are watching their brother fly the kite.	Tyrone and Sam watchin' they broth(schwa) fin' the kite. Tyrone and Sam are watching their brother find the kite.	Zero-marked auxiliary (are) Possessive (they) Derhotacized Vocalic R Zero-marked final consonant cluster
AAE GAE	Ethan and Jada was watchin' they cousin kick the ball. Ethan and Jada were watching their cousin kick the ball.	Ethan and Jada was watchin' they cousin buy the ball. Ethan and Jada were watching their cousin buy the ball.	Subject-verb-agreement (was/were) Possessive (they) Alveolarization
AAE GAE	Brianna be helpin' the boy play the piano. Every day, Brianna helps the boy play the piano.	Brianna be helpin' the boy fin' the piano. Every day, Brianna helps the boy find the piano.	Invariant be Alveolarization Zero-marked final consonant cluster
AAE GAE	They was helpin' they sist(schwa) pour the juice. They were helping their sister pour the juice.	They was helpin' they sist(schwa) fin' the juice. They were helping their sister find the juice.	Derhotacized Vocalic R Subject-verb-agreement (was/were) Alveolarization Zero-marked final consonant cluster
AAE	Caleb be helpin' his frien' Duke read the book.	Caleb be helpin' his frien' Duke choose the book.	Invariant be Alveolarization

GAE	Every day, Caleb helps his friend Duke read the book.	Every day, Caleb helps his friend Duke choose the book.	Zero-marked final consonant cluster
AAE GAE	They be watchin' they frien' Sam throw the football. Every day, they watch their friend Sam throw the football.	They be watchin' they frien' Sam choose the football. Every day, they watch their friend Sam throw the football.	Invariant be Alveolarization Zero-marked final consonant cluster
AAE GAE	They was helpin' they aunt toast the bread. They were helping their aunt toast the bread.	They was helpin' they aunt choose the bread. They were helping their aunt choose the bread.	Subject-verb-agreement (was/were) Alveolarization Possessive (they)
AAE GAE	They was helpin' they frien' win the trophy. They were helping their friend win the trophy.	They was helpin' they frien' fin' the trophy. They were helping their friend find the trophy.	Subject-verb-agreement (was/were) Alveolarization Possessive (they) Zero-marked final consonant cluster

Appendix C: Stimuli & Norming Procedures

Appendix C1. Description of the ***visual*** stimuli for noun referents displayed in the visual world paradigm in Experiment 1 and 2 and the procedures for norming all images.

Experiment 1

Photographs from publicly available databases were used for the referents in this study. All images were edited using Adobe Acrobat Photoshop so that all objects were approximately the same size, 400 pixels on the longest side. All objects were superimposed on a grey background to enhance visibility. The canvas that surrounded the edited image was 450x450 pixels.

All noun-image pairs were normed in two preschool classrooms across two separate schools. One classroom included 13 preschool children who came from families that had higher levels of maternal education (university day care center). Another classroom included 17 children whose families had lower levels of maternal education (Head Start classroom). During the norming procedure, children were shown a set of 4 images (a target referent and 3 competing referents: phonological, semantic, and unrelated). Children were instructed to select an image that best aligned with word named by a research assistant. All words that were recognized less than 80% across both classrooms were replaced and then renormed until an 80% criteria was reached.

Experiment 2

Photographs from publicly available databases were used for the referents in Experiment 2. All images were edited using Adobe Acrobat Photoshop so that all objects were approximately the same size, 400 pixels on the longest side. All objects were superimposed on a grey background to enhance visibility. The canvas that surrounded the edited image was 450x450 pixels.

All images for Experiment 2 were normed in two kindergarten classrooms at the Center for Young Children, a school located at the University of Maryland. The classrooms consisted of children ages 5 and 6 years old. All images were normed with a diverse group of children, but did not necessarily include children who were speakers of AAE. Most children were perceived to be speakers of GAE.

The procedure for image norming was the same as in Experiment 1 with the exception that instead of two images of each referent, children were shown 4 images of each referent. Display pages were created to be the same as the trial displays children would be shown in the experiment. Each trial display included a picture of the target referent and 3 unrelated items. Children were shown each trial page and were instructed to point the object that matched the word produced by the research assistant. A total of 18 children (11 5-year-olds and seven 6-year-olds) participated in norming. Images were selected if they were recognized by at least 85% of the children. Images that were below this level were replaced, but were not renormed. All images were recognized by at least 90% of the children with the exception of one of the four images of the word *horn*. The image of horn that fell below the 85% level was replaced with a different image that the researchers perceived to be similar to an image of horn that was already normed and met the criteria for inclusion.

Appendix C2. Description of the ***auditory*** stimuli norming procedures (i.e., selection of noun referents for predictive verbs) and the transformation of sentences to include AAE referents.

We used Amazon Mechanical Turk to select predictive verbs for Experiment 2. The degree to which semantically informative verbs predicted target referents was measured using a cloze task. Adults Turkers saw sentences presented orthographically with a missing final word (e.g., "Every day, Caleb helps his friend Duke read the ______"). Participants were instructed to type in the final word that best completed the sentence. The instructions prompted them to type first word that emerged in their mind while reading the sentence. A minimum of 500 adults completed this norming task.

All final words within each sentence were selected on the basis of relative predictability, age of acquisition, and the ease of imageability. In terms of predictability, the criteria for inclusion was that all referents chosen had to have a prediction proportion of at least .1. The predictive stimuli included target referents that ranged in predictability from a proportion of .125 to .799. In other words, items did range in their level of predictability, some referents being more predictive than others. The first 10 stimuli were selected on the basis of this criterion level. Of these items, the target referent included in the study was the most frequent response. For two of the items, we selected responses with either the third (n=1, *juice*) or the fifth most frequent (n=1, *trophy*). This occurred because some of the most frequent responses contained referents that would be difficult to image or were outside of the range of age of acquisition for children between 5 and 9 years old. For example, *trophy* was chosen in place of the most frequent responses including *game, race, lottery, and prize.* To ensure that there would be sufficient items for subsequent analyses, two additional predictive verbs were included but were not normed. Despite not being normed, visualization of the effects across items suggest that predictive effects were not hampered by this (See Figure XX below).

Finalizing stimuli procedure. The final set of stimuli underwent two additional processing phases. First, the preambles of the final sentences were slightly altered to introduce additional naturalistic AAE features. Some of the nouns referring to characters within the sentences were changed (e.g., *They were watching* was replaced with *David and Tyler were watching)* and some of the verbs in the preamble were changed (e.g., *They <u>help</u> their brother read* was replaced with *They <u>are helping</u> their brother read*). Second, the duration of all sentences were normed to ensure that the onset of the critical verb (*read*) and target referent (*book*) across dialect conditions were similar. This was important for measuring gaze-contingent looks throughout the duration of the eye-tracking experiment and for later comparison of looks to target image as a function of dialect familiarity and semantic predictability. Lastly, we obtained perceptual ratings of AAE auditory stimuli by 4 naïve adult listeners who self-identified as bidialectal speakers of AAE and GAE. Sentences were rated on a Likert Scale of values ranging from 0 to 5 to evaluate the listener's perception of dialect density (0 = does not sound like AAE and 5 = sounds very AAE-like). Auditory stimuli that had an average rating of 3.8 or greater were included in Experiment 2.

Appendix D: Additional figures Chapter 2 (Experiment 2)

Figure D1. Grand mean looks to the target image as a function of semantic predictability. The larger circles reflect the grand mean, the smaller points represent individual data points around that mean.

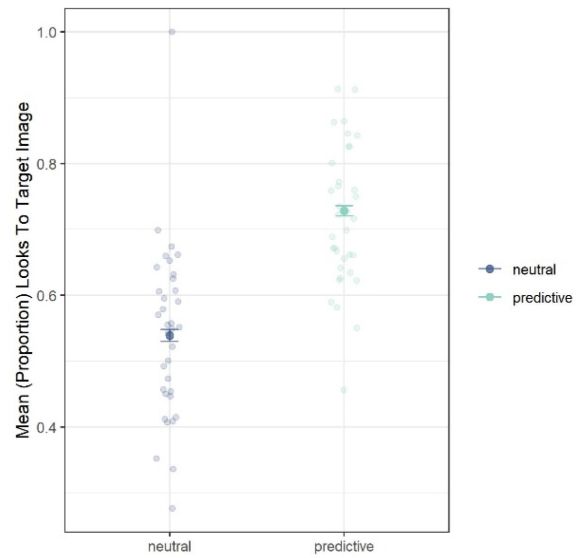
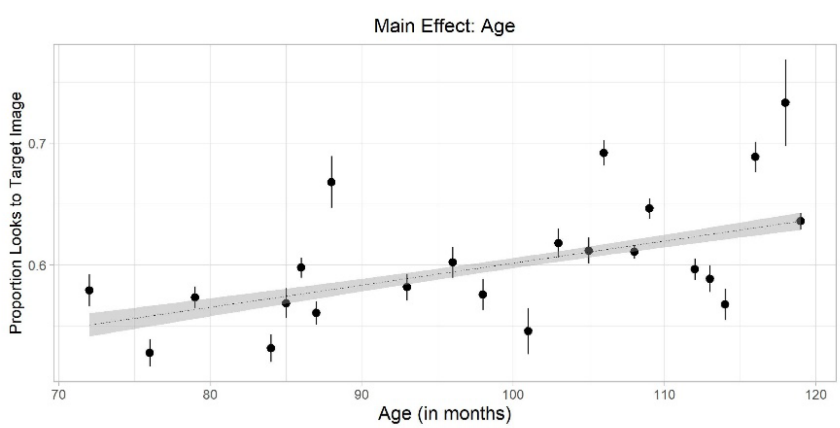


Figure D2. Relationship between age and overall sentence comprehension (i.e., the grand mean across both semantic predictability conditions). *Top panel*: The linear relationship between age (plotted as continuous) and looks to the target image. *Bottom panel:* Looks to the target image across younger and older children (groups split at the median value of age).



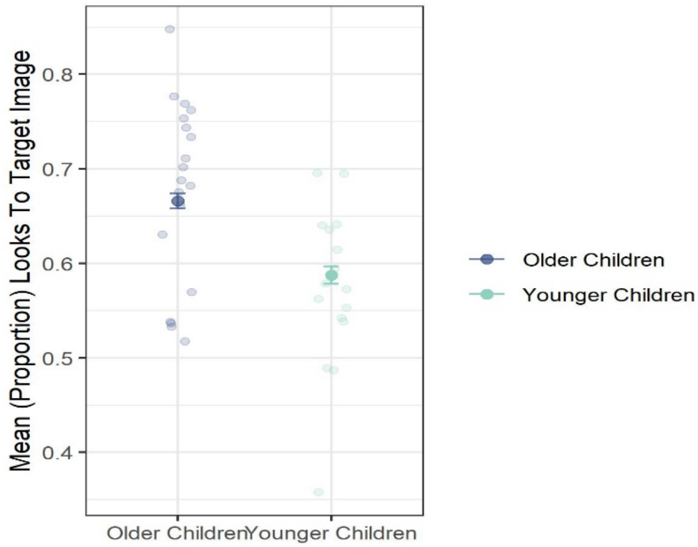


Figure D3. Looks to target image for the two Dialects Heard conditions and the two Semantic Predictability conditions in the cluster region of analysis (4500ms to 4800ms).

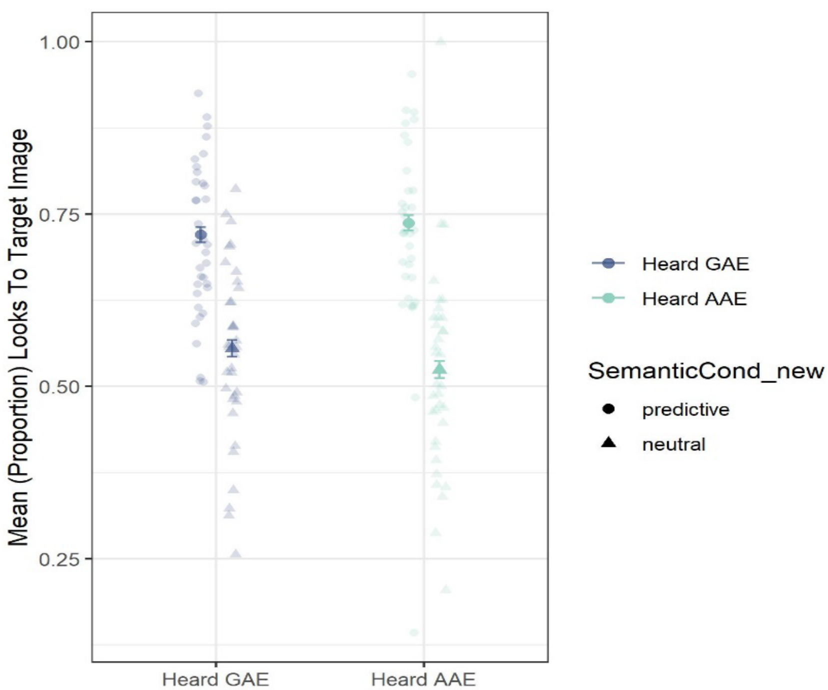
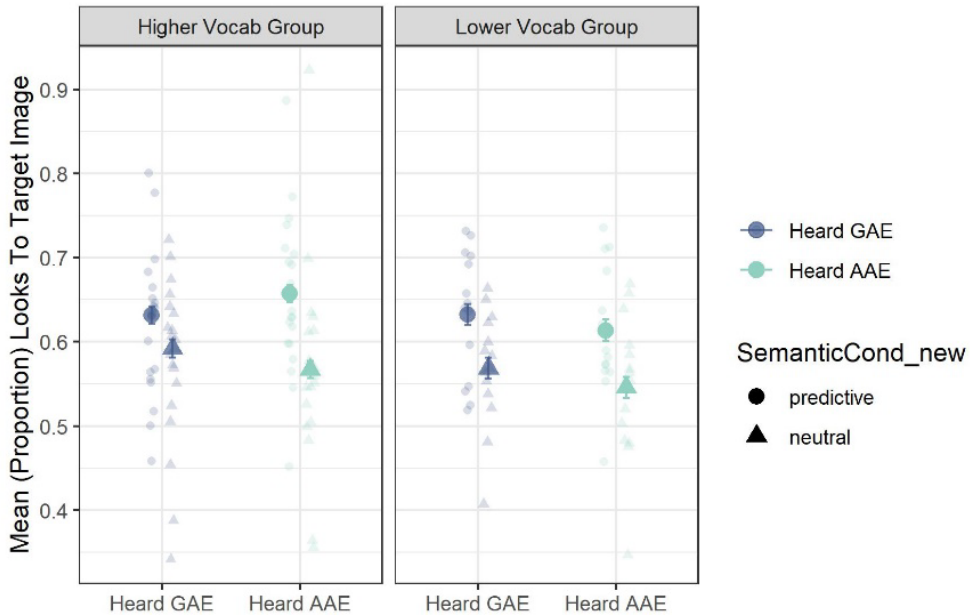


Figure D4. Looks to target image for high and low vocabulary groups (median split) for the two Dialects Heard conditions and the two Semantic Predictability conditions in the region of analysis (4500ms to 4800ms.



Appendix E: Additional model estimates for Chapter 3

Table E1.
Response accuracy among European American listeners, summary of generalized linear model estimates. The measure of vocabulary is the raw score from the NIH Toolbox PVT.

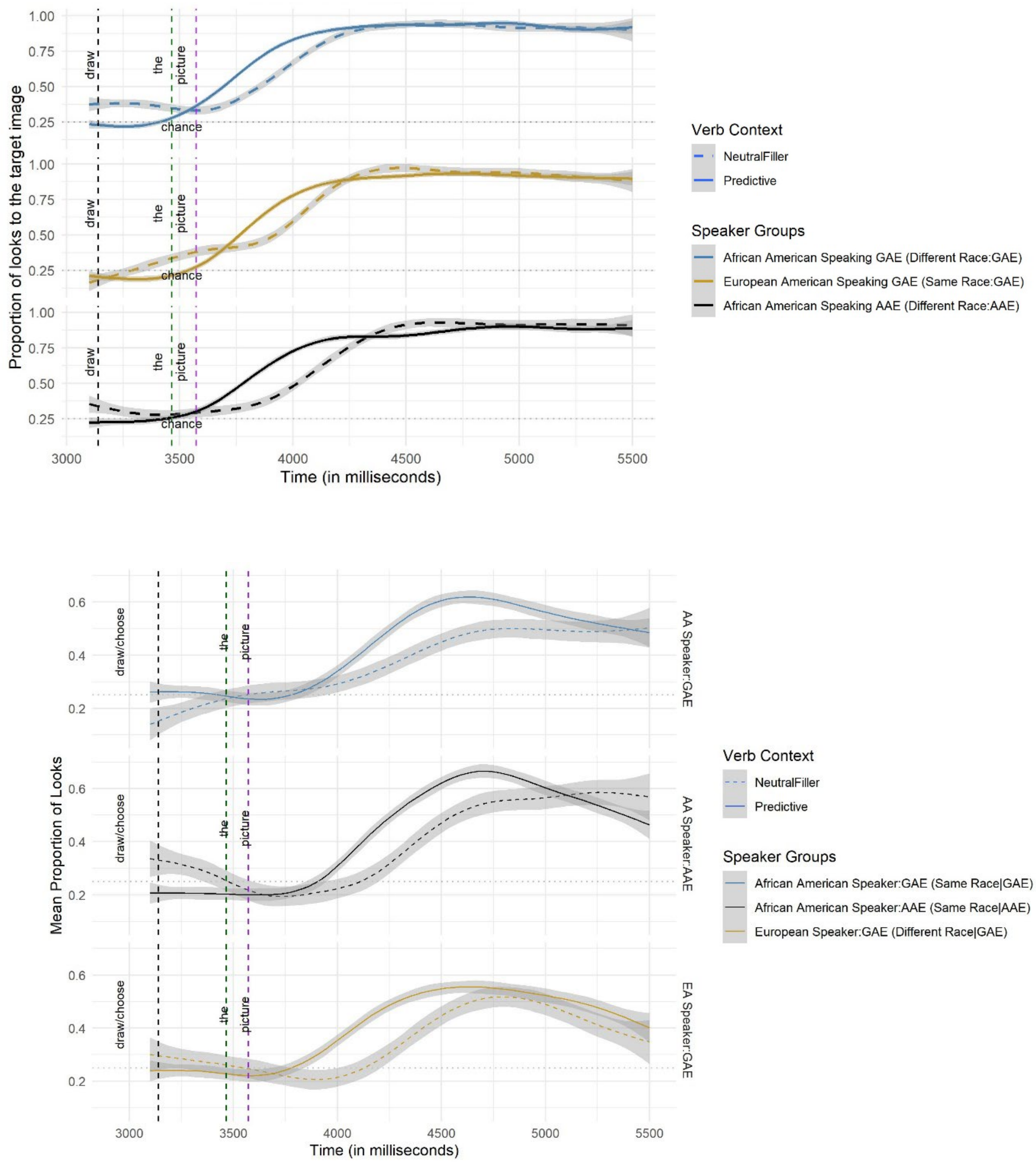
R syntax: Binary-Coded Accuracy ~ Vocabulary (centered raw PVT score) + Age (centered) + Speaker Group (dummy coded) + Speaker Group * Age + Speaker Group * Vocabulary + error

Fixed Effects	***b***	**SE**	**t**
Constant	.99	.003	304.03 **
Vocabulary	.00	.004	.15
Age	.00	.004	.89
Speaker Group (Different Race|GAE)	.00	.003	.37
Speaker Group (Different Race|AAE)	.00	.003	1.4
Speaker Group (Different Race|GAE)*Vocabulary	.00	.004	.07
Speaker Group (Different Race|AAE 2)*Vocabulary	.00	.004	.27
Speaker Group (Different Race|GAE)*Age	.00	.004	.21
Speaker Group (Different Race|AAE)*Age	.00	.004	.31

*$p < .001$***, $p < .01$ **, $p < .05$* , $p \geq .05$+*

Appendix F: Semantic Prediction Effects in Chapter 3

Figure F1. Visual illustration of semantic prediction effects replicated with fewer items for European American listeners (top plot) and African American listeners (bottom plot).



Appendix G: Stimuli List (including filler and neutral verb contexts)

Table G1. Sentences included in Experiment 3.1 and 3.2 across all verb contexts (predictive, neutral verbs, and filler items) and dialects heard (AAE and GAE).

Verb Condition	**Stimuli in AAE**	**Stimuli in GAE**	**AAE Features Included**
Predictive	They was helpin' they aunt bake the cookies	They were helping their aunt bake the cookies	subject-verb agreement possessive they
Predictive	Alyssa watchin' her brothə blow the horn	Alyssa is watchin' her brother blow the horn	alveolarization schwa zero auxiliary
Predictive	David and Tyler helpin' they sistə climb the tree	David and Tyler are helping their sister climb the tree	alveolarization possessive they schwa zero auxiliary
Predictive	Erik and Dan watchin' they frien’ draw the picture	Erik and Dan are watching their friend draw the picture	zero-marked final consonant alveolarization possessive they zero auxiliary
Predictive	They was watchin' they cousin drink the milk	They were watching their cousin drink the milk	alveolarization possessive they subject-verb agreement
Predictive	They was watchin’ they frien’ Jake drive the car	They were watching their friend Jake drive the car	subject-verb agreement alveolarization possessive they zero-marked final consonant
Predictive	Molly and Ty was watchin’ Kate eat the sandwich	Molly and Ty were watching Kate eat the sandwich	subject-verb agreement alveolarization
Predictive	Tyrone and Sam watchin’ they brothə fly the kite	Tyrone and Sam are watching their brother fly the kite	zero auxiliary possessive they alveolarization schaw
Predictive	Ethan and Jada was watchin' they cousin kick the ball	Ethan and Jada were watching their cousn kick the ball	subject-verb agreement possessive they alveolarization
Predictive	Samantha watchin' her frien' pet the cat	Samantha is watching her friend pet the cat	subject-verb agreement alveolarization zero auxiliary zero-marked final consonant
Predictive	Brianna be helpin' the boy play the piano	Every day, Brianna helps the boy play the piano	alveolarization habitual be
Predictive	They was helpin' they sistə pour the juice	They were helping their sister pour the juice	subject-verb agreement alveolarization possessive they schaw
Predictive	Caleb be helpin' his frien’ read the book	Every day, Caleb helps his friend read the book	alveolarization habitual be zero-marked final consonant

Predictive	David and Mike was watchin' they frien' ride the bike	David and Mike were watching their friend ride the bike	subject-verb agreement alveolarization possessive they zero-marked final consonant
Predictive	Caleb be helpin' his frien' Sam throw the football	Every day, Caleb helps his friend Sam throw the football	habitual be alveolarization zero-marked final consonant
Predictive	They was helpin' they grandpa toas' the bread	They were helping their grandpa toast the bread	subject-verb agreement alveolarization possessive they zero-marked final consonant
Predictive	Jayden be helpin' his grandma walk the dog	Every day, Jayden helps his grandma walk the dog	habitual be alveolarization
Predictive	They was helpin' they frien' win the trophy	They were helping their friend win the trophy	subject-verb agreement alveolarization possessive they zero-marked final consonant
Neutral	They was helpin' they grandpa buy the bread	They were helping their grandpa buy the bread	subject-verb agreement alveolarization possessive they
Neutral	They was helpin' they coach buy the trophy	They were helping their coach buy the trophy	subject-verb agreement alveolarization possessive they
Neutral	Tyrone and Sam watchin' they brothə buy the kite	Tyrone and Sam are watching their brother buy the kite	alveolarization zero auxiliary possessive they
Neutral	Molly and Ty was watchin' Kate buy the sandwich	Molly and Ty were watching Kate buy the sandwich	subject-verb agreement alveolarization
Neutral	They was watchin' they frien' Jake buy the car	They were watching their friend Jake buy the car	subject-verb agreement alveolarization possessive they zero-marked final consonant
Neutral	Ethan and Jada was watchin' they cousin buy the ball	Ethan and Jada were watching their cousn buy the ball	subject-verb agreement alveolarization possessive they
Neutral	Brianna be helpin' the boy buy the piano	Every day, Brianna helps the boy buy the piano	habitual be alveolarization
Neutral	Alyssa watchin' her brothə choose the horn	Alyssa is watching her brother choose the horn	alveolarization zero auxiliary
Neutral	Caleb be helpin' his frien' choose the book	Every day, Caleb helps his friend choose the book	habitual be alevolarization zero-marked final consonant
Neutral	They was helpin' they aunt choose the cookies	They were helping their aunt choose the cookies	subject-verb agreement alveolarization possessive they
Neutral	Samantha watchin' her frien' choose the cat	Samantha is watching her friend choose the cat	alveolarization zero auxiliary zero-marked final consonant

Neutral	They was helpin' they sistə fin’ the juice	They were helping their sister find the juice	subject-verb agreement alveolarization possessive they zero-marked final consonant
Neutral	They was watchin' they frien' fin’ the milk	They were watching their cousin find the milk	subject-verb agreement alveolarization possessive they zero-marked final consonant
Neutral	Erik and Dan watchin' they frien' find the picture	Erik and Dan are watching their friend find the picture	subject-verb agreement alveolarization possessive they zero-marked final consonant
Neutral	Jayden be helpin' his grandma fin’ the dog	Every day, Jayden helps his grandma find the dog	subject-verb agreement alveolarization zero-marked final consonant
Neutral	David and Mike was watchin' they frien' fin’ the bike	David and Mike were watching their friend find the bike	subject-verb agreement alveolarization possessive they zero-marked final consonant
Neutral	Caleb be helpin' his frien' Sam hol’ the football	Every day, Caleb helps his friend Sam hold the football	habitual be alveolarization zero-marked final consonant
Neutral	David and Tyler watchin' they sistə touch the tree	David and Tyler are watching their sister touch the tree	alveolarization zero auxiliary possessive they schwa
Filler	Bree watchin' her frien' Kate buy the shirt	Bree is watching her friend Kate buy the shirt	alveolarization zero auxiliary zero-marked final consonant
Filler	Samantha watchin' her frien' buy the crayon	Samantha is watching her friend buy the crayon	subject-verb agreement alveolarization zero auxiliary possessive they zero-marked final consonant
Filler	Lew help his brothə buy the whistle	Lew helps his brother buy the whistle	zero-marked third person singular –s
Filler	He help his uncle buy the truck	He helps his uncle buy the truck	zero-marked third person singular –s
Filler	Jasenia helpin' her aunt buy the plant	Jasenia is helping her aunt buy the plant	alveolarization zero auxiliary
Filler	Jayden watchin' his uncle choose the chair	Jayden is watching his uncle choose the chair	zero auxiliary alveolarization
Filler	Jayden watchin' his uncle choose the shoes	Jayden is watching his uncle choose the shoes	alveolarization zero auxiliary
Filler	Molly be helpin' her sistə choose the scissors	Every day, Molly helps her sister choose the scissors	habitual be alveolarization schwa

Filler	David helpin' his grandma choose the santa	David is helping his grandma choose the santa	zero auxiliary alveolarization
Filler	Bree helpin' her frien' choose the horse	Bree is helping her friend choose the horse	zero auxiliary zero-marked final consonant alveolarization
Filler	Taylor helpin' her mom choose the grapes	Taylor is helping her mom choose the grapes	zero auxiliary alveolarization
Filler	Tyler and Alex was helpin' they frien' find the bell	Tyler and Alex were helping their friend find the bell	subject-verb agreement alveolarization zero-marked final consonant possessive they
Filler	Malik helpin' his grandma fin' the dress	Malik is helping his grandma find the dress	alveolarization zero auxiliary zero-marked final consonant
Filler	Trevor watchin' his frien' fin' the mittens	Trevor is watching his friend find the mittens	zero auxiliary alveolarization zero-marked final consonant
Filler	They helpin' they sistə fin' the doll	They are helping their sister find the doll	alveolarization zero auxiliary possessive they zero-marked final consonant
Filler	Tiana helpin' the boy fin' the footprint	Tiana is helping the boy find the footprint	alveolarization zero auxiliary zero-marked final consonant cluster
Filler	Taylor help her grandpa fin' the mail	Taylor helps her grandpa find the mail	zero-marked third person singular -s zero-marked final consonant
Filler	Alyssa watch her sistə get the can	Alyssa watches her sister get the can	zero-marked third person singular –s schwa

Table G2. Distribution of speaker images by race across speaker group conditions.

Experimental List	European American (EA) GAE Speaker	African American (AA) GAE Speaker	African American (AA) AAE Speaker
List 1	EA Image A	AA Image A	AA Image C
List 2	EA Image B	AA Image C	AA Image B
List 3	EA Image C	AA Image B	AA Image A
List 4	EA Image D	AA Image A	AA Image C

Appendix H: Auditory stimuli list for Chapter 4

Table H1. *Stimuli list for Experiments 4.1 and 4.2 (Some Trials)*

Dialect Heard	Vignettes	Critical sentences
AAE	The boys and girls was at the park gettin' balls and balloons from they cousin. Daniel and Mia aksed to play with the balls, so they cousin gave each of 'em the balls. Isaac was gon play on the swing, but Jade wanted to play with the balloons.	Find the girl that have some of the balls.
GAE	The boys and girls were at the park gettin' balls and balloons from their cousin. Daniel and Mia asked to play with the balls, so their cousin gave each of them the balls. Isaac was going to play on the swing, but Jade wanted to play with the balloons.	Find the girl that has some of the balls.
AAE	The boys and girls was cleanin' up the baskets and bats on the field. Daniel and Mia was aksed to pick up the baskets by they teacher. Isaac was too young to help, so Jade had to clean up the bats.	Find the girl that have some of the baskets.
GAE	The boys and girls were cleaning up the baskets and bats on the field. Daniel and Mia were asked to pick up the baskets by their teacher. Isaac was too young to help, so Jade had to clean up the bats.	Find the girl that has some of the baskets.
AAE	The boys and girls was shopping at they school store. The store was selling stuffed bears and berries. Daniel and Mia was buyin' the stuffed bears for they sisters. Isaac ain't want nothing, but Jade bought the berries for her aunt.	Find the girl that have some of the bears.
GAE	The boys and girls were shopping at their school store. The store was selling stuffed bears and berries. Daniel and Mia were buyin' the stuffed bears for their sisters. Isaac didn't want anything, but Jade bought the berries for her aunt.	Find the girl that has some of the bears.
AAE	The boys and girls was in the garage pickin' boxes and bottles for they art project. Daniel and Mia wanted the bottles for they project. Jade been finished her project, so he let Isaac keep the boxes.	Find the boy that have some of the boxes.
GAE	The boys and girls were in the garage pickin' boxes and bottles for their art project. Daniel and Mia wanted the bottles for their art project. Jade finished her project a long time ago, so she let Isaac keep the boxes.	Find the boy that has some of the boxes.
AAE	The boys and girls was at the park chasing bunnies and butterflies with they dad. Daniel and Mia caught bunnies, but Jade was too slow and ain't catch nothin'.	Find the boy that have some of the bunnies.

	Isaac got lucky and caught the butterflies in her jar.	
GAE	The boys and girls were at the park chasing bunnies and butterflies with their dad. Daniel and Mia caught bunnies, but Jade was too slow and didn't catch anything. Isaac got lucky and caught the butterflies in his jar.	Find the boy that has some of the bunnies.
AAE	The boys and girls was at the store with they aunt buyin candles and candies. They aunt was watching' Daniel and Mia smell the candles, so she bought them candles. Jade been knew Isaac loved candy, so she let him have the candies.	Find the boy that have some of the candles.
GAE	The boys and girls were at the store with their aunt buying candles and candies. Their aunt was watching' Daniel and Mia smell the candles, so she bought them candles. Jade knew that Isaac loved candy, so she let him have the candies.	Find the boy that has some of the candles.
AAE	The boys and girls was showing they teacher what they brought for show and tell. Daniel and Mia both got new cats from they parents. Isaac was sick and didn't bring nothing to show, but Jade brought carrots from her grandma's garden.	Find the girl that have some of the cats.
GAE	The boys and girls were showing their teacher what they brought for show and tell. Daniel and Mia both got new cats from their parents. Isaac was sick and didn't bring anything to show, but Jade brought carrots from her grandma's garden.	Find the girl that has some of the cats.
AAE	The boys and girls was at the store with they aunt buyin' food for lunch. Daniel and Mia was cravin' somethin' salty, so they both pick' chips. Jade was feelin' too sick to eat, but Isaac was real hungry and wanted chicken.	Find the boy that have some of the chips
GAE	The boys and girls were at the store with their aunt buying food for lunch. Daniel and Mia were craving something salty, so they both picked chips. Jade was feeling too sick to eat, but Isaac was really hungry and wanted chicken.	Find the boy that has some of the chips.
AAE	The boys and girls was gettin' dolls and dogs for they Christmas presents. Daniel and Mia been beggin' they parents for dolls, so they got dolls for Christimas. They parents knew Jade be allergic to animals, so they ain't get her nothin'. but Isaac wanted new dogs to play with.	Find the boy that have some of the dolls.
GAE	The boys and girls were getting dolls and dogs for their Christmas presents. Daniel and Mia had been begging their parents for dolls for a really long time, so they got dolls for Christmas. Their parents already knew Jade was allergic to animals, so they didn't get her anything, but Isaac wanted new dogs to play with.	Find the boy that has some of the dolls.
AAE	The boys and girls was gettin' money and monkeys for they birthday. Daniel and Mia had aksed for monkeys like the ones they seen at the	Find the girl that have some of the monkeys.

	zoo. Isaac knew Jade been savin' up for a bike, so he let her keep his money.	
GAE	The boys and girls were gettin' money and monkeys for their birthday. Daniel and Mia asked for monkeys like the ones they had seen at the zoo. Isaac knew Jade had been saving for a bike, so he let her keep his money.	Find the girl that has some of the monkeys
AAE	The boys and girls was buyin' gifts for they aunt. They was at the store looking at pillows and pictures. Daniel and Mia BIN knew that they aunt liked pillows, so they both bought her pillows. Isaac ain't really like her, so he ain't buy her nothing. Jade felt real bad about that, so she bought pictures for they aunt and said it was from her and Isaac.	Find the girl that have some of the pictures.
GAE	The boys and girls were buyin' gifts for their aunt. They were at the store looking at pillows and pictures. Daniel and Mia already knew that their aunt liked pillows, so they both bought her pillows. Isaac didn’t really her, so he didn't buy her anything. Jade felt badly about that, so she bought pictures for their aunt and said it was from her and Isaac.	Find the girl that has some of the pictures.
AAE	The boys and girls at they grandpa's house was gettin' peas and pizza for dinner. The grandpa gave pizza to Daniel and Mia because they was real hungry. He BIN knew Jade ate healthy, but Isaac needed to eat more vegetabes.	Find the boy that have all of the pizza.
GAE	The boys and girls at their grandpa's house were gettin' peas and pizza for dinner. The grandpa gave pizza to Daniel and Mia because they were really hungry. He already knew that Jade was a healthy eater, but Isaac needed to eat more vegetables.	Find the boy that has some of the pizza.
AAE	The boys and girls was at they cousin's Christmas party. At the end of the party, they cousin was givin' away presents and pretzels. Daniel and Mia aksed for presents to give to they friends at school who ain't get any presents. Isaac had lef' the party too early to get anything, but Jade aksed to take the pretzels home.	Find the girl that have some of the pretzels.
GAE	The boys and girls were at their cousin's Christmas party. At the end of the party, their cousin was giving away presents and pretzels. Daniel and Mia asked for presents to give to their friends at school who didn't get anything for Christmas. Isaac left the party too early to get anything, but Jade asked to take the pretzels home.	Find the girl that has some of presents.
AAE	The boys and girls at the park was pettin' rabbits and rats from the nature center. Daniel and Mia was bof excited to see the rabbits, so they was	Find the boy that have some of the rabbits

	pettin' the rabbits. Jade been scared of rats since the day she was born so she ain't pet none, but Isaac enjoy hol'in' 'em for fun.	
GAE	The boys and girls at the park were pettin' rabbits and rats from the nature center. Daniel and Mia were both excited to see the rabbits, so they were pettin' the rabbits.Jade had been scared of rats since the day she was born, so she didn't pet any of them, but Isaac enjoyed holding them for fun.	Find the boy that has some of the rabbits.
AAE	The boys and girls was at the store buyin' robes and roses for Mother's day. Daniel and Mia went to the flower shop and bought roses for they mom. Jade was plannin' to cook for her mom so she ain't buy her nothing at the store. Isaac's mom stay shoppin' for clothes, so he bought his mom pretty robes.	Find the boy that have some of the roses.
GAE	The boys and girls were at the store buyin' robes and roses for Mother's day. Daniel and Mia went to the flower shop and bought roses for their mom. Jade was plannin' to cook for her mom so she didn't buy her anything at the store. Isaac's mom shopped for clothes often, so he bought his mom pretty robes.	Find the boy that has some of the roses.
AAE	The boys and girls at the beach was getting sandals and sandwiches from they mom. Daniel and Mia was gon' walk around the beach, so they mom gave them sandals. Isaac was gon' keep swimmin', but Jade was ready to eat the sandwiches.	Find the girl that have some of the sandals.
GAE	The boys and girls at the beach were getting sandals and sandwiches from their mom. Daniel and Mia were going to walk around the beach, so their mom gave them sandals. Isaac was going to keep swimmin', but Jade was ready to eat the sandwiches.	Find the girl that has some of the sandals.
AAE	The boys and girls on the soccer team was gettin' socks and soccer balls. Daniel and Mia ain't feel like practicing, so they coach gave them socks. Everyone been knew Jade was a good soccer play who ain't need no practice. But the coach knew Isaac needed more practice, so he gave him the soccer balls.	Find the boy that have some of the socks.
GAE	The boys and girls on the soccer team were gettin' socks and soccer balls. Daniel and Mia did not feel like practicing, so their coach gave them socks. Everyone already knew Jade was a good soccer play who didn't need any practice. But the coach knew Isaac needed more practice, so he gave him the soccer balls.	Find the boy that has some of the socks.
AAE	The boys and girls was on a scavenger hunt lookin' for ties and tigers at the zoo. Daniel and Mia foun' the ties on the floor. Jade wasn't paying	Find the boy that have some of the ties.

	attention so she ain't fin' nothing, but Isaac foun' the tigers.	
GAE	The boys and girls were on a scavenger hunt lookin' for ties and tigers at the zoo. Daniel and Mia found the ties on the floor. Jade wasn't paying attention so she didin't find anything, but Isaac found the tigers.	Find the boy that has some of the ties.
AAE	The boys and girls at the zoo was gettin' turtles and turkeys to hol' from the zookeeper. Daniel and Mia was scared of the turkeys, so they got to hol' the turtles. Isaac been held all the animals at the zoo, so he let Jade hol' the turkeys.	Find the girl that have some of the turtles.
GAE	The boys and girls at the zoo were gettin' turtles and turkeys to hold from the zookeeper. Daniel and Mia were scared of the turkeys, so they got to hold the turtles. Isaac has already held all of the animals at the zoo, so he let Jade hold the turkeys.	Find the girl that has some of the turtles.
AAE	The boys and girls at the kitchen table was gettin' watermelons and waffles from they grandma. Daniel and Mia was allergic to watermelon, so they grandma gave them waffles. Isaac ain't want nothing to eat, but they grandma knew Jade needed to eat more fruits.	Find the girl that have some of the waffles.
GAE	The boys and girls at the kitchen table were gettin' watermelons and waffles from their grandma. Daniel and Mia were allergic to watermelon, so their grandma gave them waffles. Isaac didn't want anything to eat, but their grandma knew Jade needed to eat more fruits.	Find the girl that has some of waffles.

Table H2. *Stimuli Lists for Experiments 4.1 and 4.2 (All-Filler Trials)*

Story Vignette in AAE	***All* trial sentence in AAE**	**Story Vignette in GAE**	***All* trial sentence in GAE**
The boys and girls was at they grandmpa's house gettin' apples and cookies. Daniel and Mia was allergic to applies, so they grandma gave them cookies. Isaac was feelin' too sick to eat, so he gave his apples to Jade.	Find the girl that have all of the apples.	The boys and girls were at their grandma's house getting apples and cookies. Daniel and Mia were allergic to apples, so their grandma gave them cookies. Isaac was feeling too sick to eat, so he gave his apples to Jade.	Find the girl that has all of the apples.
The boys and girls was at the farmer's market getting bananas and carrots. Daniel and Mia wanted vegetables, so they aksed they parents to get them carrots. Isaac ain't see nothing he wanted to buy, but Jade was excited to taste the bananas.	Find the girl that have all of the bananas.	The boys and girls were at the farmer's market getting bananas and carrots. Daniel and Mia wanted vegetables, so they asked their parents to get them carrots. Isaac didn't see anything he wanted to buy, but Jade was excited to taste the bananas.	Find the girl that has all of the bananas.
The boys and girls was in they classroom gettin' bells and whistles from they teacher. Daniel and Mia was scared of the bells, so they aksed they teacher for the whistles. Jade aint get none because she was at home sick, so Isaac got the bells.	Find the boy that have all of the bells.	The boys and girls were in their classroom getting bells and whistles from their teacher. Daniel and Mia were scared of the bells, so they asked their teacher for the whistles. Jade didn't get any because she was at home sick. So Isaac was given the bells.	Find the boy that has all of the bells
The boys and girls was at the nature center learning 'bout bugs and bunnies. Daniel and Mia was interested in learning about the bunnies, so they teacher let them hol' the bunnies. Jade missed the trip because she was sick, so Isaac got to hol' the bugs by hisself.	Find the boy that have all of the bugs	The boys and girls were at the nature center learning about bugs and bunnies. Daniel and Mia were interested in learning about the bunnies, so their teacher let them hold the bunnies. Jade missed the trip because she was sick, so Isaac got to hold the bugs by himself.	Find the boy that has all of the bugs
The boys and girls was at the library on a scavenger hunt looking for couches and tables. Daniel and Mia ain't find any couches, but they each foun' tables. Isaac ain't bring his glasses so he couldn't see nothin' , but Jade foun' the couches.	Find the girl that have all of the couches.	The boys and girls were at the library on a scavenger hunt looking for couches and tables. Daniel and Mia didn't find any couches, but they each found tables. Isaac didn't bring his glasses so he couldn't see anything, but Jade found the couches.	Find the girl that has all of the couches.
The boys and girls was at the farm pettin' horses and cows. Daniel and Mia BIN ridin' horse since they was young, so they was pettin' the horses. Jade was too busying lookin' at	Find the boy that have all of the cows	The boys and girls were at the farm petting horses and cows. Daniel and Mia had been horse riding since they were young, so they were petting the horses. Jade was too busy looking at the	Find the boy that has all of the cows.

the tractors so she ain't pet no animals. Isaac was interested in the cows, so he spent the whole day pettin' the cows.		tractors, so she didn't pet any animals. Isaac was interested in the cows, so he spent the whole day pettin' the cows.	
The boys and girls was shopping at they school store. The store was selling grapes and milk. Daniel and Mia was buying milk for they younger sisters. Jade ain't want nothin' but Isaac was real hungry so he bought grapes for hisself.	Find the boy that have all of the grapes	The boys and girls were shopping at their school store. The store was selling grapes and milk. Daniel and Mia were buyin' the milk for their younger sisters. Jade didn't want anything, but Isaac was really hungry so he bought grapes for himself.	Find the boy that has all of the grapes
The boys and girls was in they music class pickin' horns and pianos. Daniel and Mia was interested in playin the pianos, so they pick the pianos. Jade ain't want to play none of the instruments, so she let Isaac have the horns.	Find the boy that have all of the horns.	The boys and girls were in their music class picking horns and pianos. Daniel and Mia were interested in playing the pianos, so they picked the pianos. Jade didn't want to play any of the instruments, so she let Isaac have the horns.	Find the boy that has all of the horns.
The boys and girls was at the mall buyin' jewels and juice. Daniel and Mia needed new jewelry for a party, so they bought the jewels. Jade ain't care about the juice or the jewels, so she ain't buy nothin'. Isaac was real thirsty while at the mall, so he bought the fruit juices.	Find the boy that have all of the juices.	The boys and girls were at the mall buying jewels and juice. Daniel and Mia needed new jewelry for a party, so they bought the jewels. Jade didn't care about the juice or jewels, so she didin't buy anything. Isaac was really thirsty while at the mall, so he bought the fruit juices.	Find the boy that has all of the juices.
The boys and girls was at the park playing wit kites and footballs. Daniel and Mia was playign wit' the footballs wit' they cousins. Isaac was too busy starin' at the clouds tryin' to fin' different shapes, but Jade was flyin' the kites.	Find the girl that have all of the kites.	The boys and girls were at the park playing with kites and footballs. Daniel and Mia were playing with the footballs with their cousins. Isaac was too busy staring at the clouds trying to find different shapes, but Jade was flyin' the kites	Find the girl that has all of the kites.
The boys and girls was at the store with they parents buyin' laptops and toy ladders. Daniel and Mia parents was buyin' them new laptops for school. Isaac ain't need none of this, but Jade wanted ladders for her dollhouse.	Find the girl that have all of the ladders.	The boys and girls were at the store with their parents buying laptops and toy ladders. Daniel and Mia's parents were buying them new laptops for school. Isaac didn't need any of these, but Jade wanted ladders for her dollhouse.	Find the girl that has all of the ladders.
The boys and girls was that the grocery store buyin' lettuce and lemons with they cousin.	Find the girl that have all of the lemons.	The boys and girls were at the grocery store buying lettuce and lemons with their cousin.	Find the girl that has all of the lemons.

Daniel and Mia wanted to eat sandwiches for they lunch, so they bought the lettuce. Isaac BIN ate lunch, but Jade wanted to make a lemon cake.		Daniel and Mia wanted to eat sandwiches for their lucnh, so they bought the lettuce. Isaac had already eaten a big lunch, but Jade wanted to make a lemon cake.	
The boys and girls was at they grandpa's house helpin' him carry the mail and watermelons. Daniel and Mia was older and much stronger, so they carried the watermelons. Jade was too young to carry nothing, but Isaac was old enough to carry the mail.	Find the boy that have all of the mail	The boys and girls were at their grandpa's house helping him carry the mail and watermelons. Daniel and Mia were older and much stronger, so they carried the watermelons. Jade was too young to carry anything, but Isaac was old enough to carry the mail.	Find the boy that has all of the mail
The boys and girls was in they classroom playin' wit' markers and marbles. Daniel and Mia like drawing, so they pick the markers. Isaac ain't want to play with none of the markers or marbles, so he let Jade have the marbles.	Find the girl that have all of the marbles.	The boys and girls were in their classroom playing with markers and marbles. Daniel and Mia liked drawing, so they picked the markers. Isaac didn't want to play with any of the markers or marbles, so he let Jade have the marbles.	Find the girl that has all of the marbles.
The boys and girls was in they aunt's kitchen gettin pans and bowls. Daniel and Mia was going to help they aunt mix everything, so they grab the bowls. Jade wasn't tall enough to reach the stove to help cook, so Isaac grab the pans to help his aunt cook.	Find the boy that have all of the pans.	The boys and girls were in their aunt's kitchen getting pans and bowls. Daniel and Mia were going to help their aunt mix everything, so they grabbed the bowls. Jade wasn't tall enough to reach the stove to help cook, so Isaac grabbed the pans to help his aunt cook.	Find the boy that has all of the pans
The boys and girls was at the store buyin' pickles and pictures. Daniel and Mia really wanted picture to hang on they bedroom walls, so they parents bought them the pictures. Jade ain't want no pictures and ain't like pickles, so she let Isaac have her pickles.	Find the boy that have all of the pickles	The boys and girls were at the store buying pickles and pictures. Daniel and Mia really wanted pictures to hang on their bedroom walls, so their parents bought them pictures. Jade didn't want any pictures and didn't like pickles, so she let Isaac have the pickles.	Find the boy that has all of the pickles
The boys an girls was at the store with they parents buying chairs and plants Daniel and Mia was hopin' to get new chairs for they bebdrooms, so they parents bought them chairs. Isaac ain't need nothin' new, but Jade really liked flowers so her parents bought her the plants.	Find the girl that have all of the plants	The boys and girls were at the store with their parents buying chairs and plants Daniel and Mia were hoping to get new chairs for their bedrooms, so their parents bought them chairs. Isaac didn't need anything new, but Jade really liked flowers so her parents bought her the plants for her room.	Find the girl that has all of the plants
The boys and girls was in the kitchen gettin' forks and spoon to eat they food.	Find the girl that have all of the spoons	The boys and girls were in the kitchen getting forks and spoons to eat their food.	Find the girl that has all of the spoons

Daniel and Mia was eatin' steak and needed to cut it with forks, so they pick the forks. Isaac BIN ate his food, but Jade was still eatin' her jello, so she pick the spoons.		Daniel and Mia were eating steak and needed to cut it with forks, so they picked the forks. Isaac had already finished eating his food, but Jade still had to eat her jello, so she picked the spoons.	
The boys and girls was at the dentist gettin' toofbrushes and toofpaste at the end of they visit. Daniel and Mia ain't need toothbrushes, so they aksed they dentist for toofpaste Jade only liked her toothbrush that was at home, so she ain't get nothin'. Isaac really liked gettin' new toothbrushes, so he aksed for the toothbrushes.	Find the boy that have all of the toothbrushes	The boys and girls were at the dentist gettin' toothbrushes and toothpaste at the end of their visit. Daniel and Mia didn't need toothtbrushes, so they asked their dentist for toothpaste. Jade only liked her toothbrush that was at home, so she didn't get anything. Isaac really liked getting new toothbrushes, so he asked for the toothbrushes.	Find the boy that has all of the toothbrushes.
The boys and girls was at the toy store with they parents gettin' trophies and trucks. Daniel and Mia was buyin' trophies for they cousins. Isaac had fell asleep on the ride to the store and ain't get nothin' Jade bought the trucks because they looked like ones she seen at her aunt's house.	Find the girl that have all of the trucks	The boys and girls were at the toy store with their parents getting trophies and trucks. Daniel and Mia were buying trophies for their cousins. Isaac fell asleep on the ride to the store and didn't get anything. Jade bought the trucks because they looked like the one she'd seen at her aunt's house.	Find the girl that has all of the trucks.

Table H3. Distribution of speaker images by race across speaker group conditions.

Experimental List	**European American (EA) GAE Speaker**	**African American (AA) GAE Speaker**	**African American (AA) AAE Speaker**
List 1	EA Image A	AA Image B	AA Image A
List 2	EA Image B	AA Image C	AA Image A
List 3	EA Image C	AA Image A	AA Image C
List 4	EA Image D	AA Image C	AA Image B

Appendix I: Additional model estimates for Chapter 4

Table I1

The generalized linear model for response accuracy, binary coded (0,1) among European American Listeners (n=20) in Experiment 4.1.

R Syntax: glm (Accuracy) ~ Vocabulary (*centered raw score*) + Age (*centered*) + Speaker Group (*dummy coded*) + error

Fixed Effects			
	b	**SE**	**z**
Constant	.98	.02	45.34 ***
Speaker Group (Same Race|GAE dialect)	-.02	.03	-.79
Speaker Group (Different Race|AAE dialect)	-.01	.03	-.38
Age (*In Months*)	-.01	.01	-1.17
Vocabulary (PVT NIH Toolbox Raw Score)	.02	.01	1.49

p <.001***

Appendix J: Additional Stimuli Details and Norming Procedures

Administration of webcam eye-tracking task. Children participated in this study either in person or remotely. For remote participation, caregivers were encouraged to have two electronic devices, a computer and an additional device such as a phone, computer, or tablet which could be used to open a Zoom link. Two devices were necessary so that children could complete the eye-tracking task on a computer and have a second device where the experimenter could be present to provide technical support or reinforcement about the child's posture or attention to the experiment. The day before a scheduled visit, parents received an email with instructions about how to join the experiment Zoom session and a link to the experiment. At the start of the virtual session, parents provided informed consent and used the instructions to set up the eye-tracking experiment with support of the examiner on the Zoom call. After the experiment was set up, children were seated in front of the computer close enough to the computer so that the webcam could record their eye movements in closer proximity. The entire experimental procedure took approximately 1 hour. Preparing the experiment took about 20 minutes, the visual world paradigm took 20 minutes, and the remaining time was dedicated to breaks and completing two additional language measures (i.e., vocabulary and language variation screeners). This procedure was used for a total of six children, but only two of the participants who participated remotely were included in the final sample. The exclusion of participants was due to equipment failure (i.e., some children had computers with malfunctioning speakers which created a noisy listening environment) or difficulties with compliance.

In-person visits took place at the University. To create environments that were similar to virtual visits, the same experimental procedures were used for the in-person visits. Informed consent was completed at the start of the session with the child's caregiver. Children were seated in a quiet room in front of a computer and positioned close to the camera. However, unlike virtual sessions, in-person visits required a single device because the experimenter was present to assist with setting up the task and providing additional instructions. The computers used during the in-person visits varied because some of the experimenters (i.e., trained research assistants in the lab) used their personal device and others used desktops in the lab if they were available. While in-person visits were not the same as remote visits at home, these steps were taken to minimize the differences across administration procedures.

Appendix K: Webcam Eye-tracking Coding Training and Calibration Procedures

During guided coding, trainees met in small groups with a lead coder and as a group the lead coder and trainees coded 6 example trials with an answer key visually available. This was done to ensure that team members learned about the application of specific codes and how different looking patterns were mapped to specific codes. During these meetings, the lead coder provided additional guidance and feedback as often as necessary. In the second phase of training, trainees were provided with a new set of example trials which they coded independently (i.e., without a lead coder) but with the use of an answer sheet. This procedure provided opportunities for coders to engage in independent learning and practice of skills acquired in the first few meetings with the lead coder. Questions related to trials coded during this phase were addressed in a separate group meeting with the lead coder. In the last training phase, coders were provided another set of trials which they coded independently, but this time without access to an answer key. This step was crucial for providing trainees with opportunities to practice coding with less support. During this stage, trainees coded videos and then turned in their responses to the lead coder who would evaluate their reliability against an answer key and provide additional feedback and guidance as deemed necessary. The reliability of coders was also examined for inter-rater reliability against an answer sheet. Inter-rater reliability was calculated using a default function provided by Peyecoder. In depth details about the interrater reliability calculator can be found (here).

In this study, we used the ‘Compare against” function which calculates three specific measures: frame agreement, comparable trials, and shift agreement. Frame agreement corresponds to the general agreement between coders about gaze location (i.e., the percentage of frames in which two or more people coded the same response). This index provides a measure of consistency or general agreement about where children are looking on the screen. Comparable trials indexes whether manual coders have coded the same number of responses. Lastly, shift agreement evaluates the percentage of agreement across time stamps of events coded with the same response within close proximity of each other. That is, it examines the extent to which coders are consistent in how the number of frames that a response and whether the timing of this response is in close proximity. A coder was deemed reliable if they demonstrated 80% agreement in at least 2 of 3 of the measures. This criterion was used because the "Compare against" function was originally designed to be used with looking while listening paradigms, in which gaze patterns are monitored in only 2 to 3 areas of interests (usually, left, right, and center for fixation points) (Pomper et al. 2022, Venker et al. 2020). In the eye-tracking task in this study, we monitored four relevant areas of interests (top-left, top-right, bottom-left and bottom-right) on screens with variable sizes and dimensions. Usually, in task with more than two area of interests, looks can be more difficult to assess, especially gaze movements in the up and down position. As a result, we used a criterion level for interrater reliability that is lower than previous studies.

Nevertheless, if trainees reached the criterion level of 80% interrater reliability in at least 2 measures (against the answer sheet), they were advanced to the calibration phase. If trainees did not reach this criterion level, they were assigned additional trials until the necessary agreement threshold was reached. The training process with four coders took approximately 5 months to complete.

Calibration and post calibration review. During the calibration process, trainees were assigned 20 calibration files across two participants from two different eye-tracking experiments. Trial assignments were based on a set of files that had been previously coded by a group of trained coders. Additionally, answer keys were developed for each of these files and were used to evaluate the calibration files completed by the trainees.

For calibration files, trainees independently coded all assigned trials. These trials were later assessed for reliability by the lead coder using the "Compare against" function. If coders reached the predetermined 80% threshold in two of three measures, they were assigned critical trials from the experiment reported in this paper. If coders did not reach the predetermined criterion, they were assigned a new set of previously coded trials. Trainees remained within the calibration phase until they met the criterion threshold. All coders met the criteria for inter rater reliability within two attempts. After coders completed training and calibration and coders advanced to independently coding videos, the lead coder periodically reviewed files to examine the quality of independently coded data. To ensure the quality of data, the lead coder would randomly evaluate an early and middle portion of trials from videos that were in progress of being coded. If glaring issues were observed, the assigned coder was asked to recode the video. If no concerns were noted by the lead coder, the coder assigned to the video continued to code the video until it was completed. This was usually done early on to avoid recoding of entire participant.

Bibliography

Aboud, F. E., Mendelson, M. J., & Purdy, K. T. (2003). Cross-race peer relations and friendship quality. *International Journal of Behavioral Development*, *27*(2), 165–173. https://doi.org/10.1080/01650250244000164

Adank, P., & McQueen, J.M. (2007). The effect of an unfamiliar region accent on the speed of word processing. *Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken, Germany,* 1925-1928.

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520–529. https://doi.org/10.1037/a0013552

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., & Lee, K. (2009). Categorization, categorical perception, and asymmetry in infants' representation of face race: Infants' representation of face race. *Developmental Science*, *13*(4), 553–564. https://doi.org/10.1111/j.1467-7687.2009.00900.x

Arnold, J.E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes, 22*(4), 537-565. DOI: 10.1080/01690960600845950

Babel, M. (2022). Adaptation to Social-Linguistic Associations in Audio-Visual Speech. *Brain Sciences*, *12*(7), 845. https://doi.org/10.3390/brainsci12070845

Babel, M., & Mellesmoen, G. (2019). Perceptual adaptation to stereotyped accents in audio-visual speech. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019,* eds. S. Calhoun, P. Escudero, M. Tabain, & P. Warren: 1004-1048.

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, *137*(5), 2823–2833. https://doi.org/10.1121/1.4919317

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL142-EL180.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and Nurture in Own-Race Face Processing. *Psychological Science*, *17*(2), 159–163. https://doi.org/10.1111/j.1467-9280.2006.01679.x

Barker, B. A., & Turner, L. M. (2015). Influences of foreign accent on preschoolers' word recognition and story comprehension. *Applied Psycholinguistics*, *36*(5), 1111–1132. https://doi.org/10.1017/S0142716414000058

Barry-Anwar, R. A., Burris, J. L., Graf Estes, K., & Rivera, S. M. (2017). Caregivers and strangers: The influence of familiarity on gaze following and learning. *Infant Behavior and Development*, *46*, 46–58. https://doi.org/10.1016/j.infbeh.2016.11.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). lme4*: linear mixed-effects models using Eigen and S4.* R package version 1.1-11.

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

Beyer, T., & Hudson Kam, C. L. (2012). First and second graders' interpretation of Standard American English morphology across varieties of English. *First Language*, *32*(3), 365–

384. https://doi.org/10.1177/0142723711427618

Blair, S., & Gya, G. (2010). The two faces of civilian in civilian crisis management: an opportunity to bring them together in the Lisbon era. *Studia Diplomatica, 63*, 105-131.

Blomquist, C., Newman, R. S., Huang, Y. T., & Edwards, J. (2021). Children With Cochlear Implants Use Semantic Prediction to Facilitate Spoken Word Recognition. *Journal of Speech, Language, and Hearing Research*, *64*(5), 1636–1649. https://doi.org/10.1044/2021_JSLHR-20-00319

Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.02, retrieved 29 November 2022 from http://www.praat.org/

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417-436.

Bott, L., & Noveck, I. A. (2004) Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*(3), 437-457.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707-729.

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440. https://doi.org/10.1016/j.cognition.2012.11.012

Bricker, P. D., & Pruzansky, S. (1966). Effects of Stimulus Content and Duration on Talker Identification. *The Journal of the Acoustical Society of America*, *40*(6), 1441–1449. https://doi.org/10.1121/1.1910246

Brungart, D. S., Barret, M. E., Cohen, J. I., Fodor, C., Yancey, C.M., & Gordon-Salant, S.

(2020). Objective assessment of speech intelligibility in crowded public spaces. *Ear and Hearing, 41,* p.68S.

Bühler, J. C., von Oertzen, T., McBride, C. A., Stoll, S., & Maurer, U. (2018). Influence of dialect use on early reading and spelling acquisition in German-speaking children in Grade 1. *Journal of Cognitive Psychology, 30*(3), 336–360. https://doi.org/10.1080/20445911.2018.1444614

Byrd, A. S., Huang, Y.T., and Edwards, J. (2022, November 3-5). *Examining how African American English- and Mainstream American English-speaking children process subject-verb agreement in sentences* [Poster Presentation]. Boston University Conference on Language Development, Boston, MA, United States.

Carminati, M. N., & Knoeferle, P. (2013). Effects of Speaker Emotional Facial Expression and Listener Age on Incremental Sentence Processing. *PLoS ONE*, *8*(9), e72559. https://doi.org/10.1371/journal.pone.0072559

Casasanto, L. S. (2008). Does Social Information Influence Sentence Processing? *Proceedings of the 30th Annual Meeting of the Cognitive Science Society,* Washington, D.C.

Charity, A. H., Scarborough, H. S., & Griffin, D. M. (2004). Familiarity With School English in African American Children and Its Relation to Early Reading Achievement. *Child Development*, *75*(5), 1340–1356. https://doi.org/10.1111/j.1467-8624.2004.00744.x

Chauncey, K., Grainger, J., & Holcomb, P. (2008). Code-switching effects in bilingual word recognition: A masked priming study with event-related potentials. *Brain and Language, 105,* 161-174.

Clement, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language, 19*(4), 360-379. https://doi.org/10.1111/j.0268-1064.2004.00263.x

Clopper, C. G. (2008). Speech processing and dialect variation in the American Midwest. *The Journal of the Acoustical Society of America*, *123*(5), 3068-3068.

Clopper, C. G. (2012). Effects of dialect variation on the semantic predictability benefit. *Language and Cognitive Processes*, *27*(7–8), 1002–1020. https://doi.org/10.1080/01690965.2011.558779

Clopper, C. G., & Bradlow, A. R. (2008). Perception of Dialect Variation in Noise: Intelligibility and Classification. *Language and Speech*, *51*(3), 175–198. https://doi.org/10.1177/0023830908098539

Connor, C. M. (2008). Language and literacy connections for children who are African American. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations,* 15, 43–53.

Connor, C. M., & Craig, H. K. (2006). African American Preschoolers' Language, Emergent Literacy Skills, and Use of African American English: A Complex Relation. *Journal of Speech, Language, and Hearing Research*, *49*(4), 771–792. https://doi.org/10.1044/1092-4388(2006/055)

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going With the Flow: Preschoolers Prefer Nondissenters as Informants. *Psychological Science*, *20*(3), 372–377. https://doi.org/10.1111/j.1467-9280.2009.02291.x

Corriveau, K. H., Meints, K., & Harris, P. L. (2009). Early tracking of informant accuracy and inaccuracy. *British Journal of Developmental Psychology*, *27*(2), 331–342. https://doi.org/10.1348/026151008X310229

Corriveau, K., & Harris, P. L. (2009a). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, *12*(1), 188–193.

https://doi.org/10.1111/j.1467-7687.2008.00763.x

Corriveau, K., & Harris, P. L. (2009b). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*(3), 426–437. https://doi.org/10.1111/j.1467-7687.2008.00792.x

Craig, H. K., Kolenic, G. E., & Hensel, S. L. (2014). African American English–Speaking Students: A Longitudinal Examination of Style Shifting From Kindergarten Through Second Grade. *Journal of Speech, Language, and Hearing Research*, *57*(1), 143–157. https://doi.org/10.1044/1092-4388(2013/12-0157)

Craig, H. K., & Washington, J. A. (2004). Grade-Related Changes in the Production of African American English. *Journal of Speech, Language, and Hearing Research*, *47*(2), 450–463. https://doi.org/10.1044/1092-4388(2004/036)

Creel, S.C. (2012). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3-5-year olds. *Developmental Science, 15*(5), 697-713.

Creel, S. C., & Bregman, M. R. (2011). How Talker Identity Relates to Language Processing: Talkers and Language Processing. *Language and Linguistics Compass*, *5*(5), 190–204. https://doi.org/10.1111/j.1749-818X.2011.00276.x

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic Processing of Accented Speech Across the Lifespan. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00479

Davis, M. J. (2010). Contrast coding in multiple regression analysis: strengths, weaknesses, and utility of popular coding structure. *Journal of Data Science, 8,* 61-73.

De Villiers, J. G., & Johnson, V. E. (2007). The information in third-person /s/: Acquisition across dialects of American English. *Journal of Child Language*, *34*(1), 133–158.

https://doi.org/10.1017/S0305000906007768

Drager, K. (2011). Speaker Age and Vowel Perception. *Language and Speech*, *54*(1), 99–121. https://doi.org/10.1177/0023830910388017

Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test (4th ed.). Minneapolis, MN: NCS Pearson.

Dunn, D. M. (2019). Peabody Picture Vocabulary Test (5th ed.). Bloomington, MN: NCS Pearson.

Edwards, J., Gross, M., Chen, J., MacDonald, M. C., Kaplan, D., Brown, M., & Seidenberg, M. S. (2014). Dialect Awareness and Lexical Comprehension of Mainstream American English in African American English–Speaking Children. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1883–1895. https://doi.org/10.1044/2014_JSLHR-L-13-0228

Erickson, L. C., & Newman, R. S. (2017). Influences of Background Noise on Infants and Children. *Current Directions in Psychological Science*, *26*(5), 451–457. https://doi.org/10.1177/0963721417709087

Erskine, M.E. (2022a). *Isolating effects of perceptual analysis and sociocultural context on children's comprehension of two dialects of English, African American English and General American English.* [Doctoral Dissertation, University of Maryland]. ProQuest Databases and Theses Global.

Erskine, M.E. (2022b). *Isolating effects of perceptual analysis and sociocultural context on children's comprehension of two dialects of English, African American English and General American English.* [Doctoral Dissertation, University of Maryland]. ProQuest Databases and Theses Global.

Erskine, M.E. (2022c). *Isolating effects of perceptual analysis and sociocultural context on children's comprehension of two dialects of English, African American English and General American English.* [Doctoral Dissertation, University of Maryland]. ProQuest Databases and Theses Global.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

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*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid Gains in Speed of Verbal Processing by Infants in the 2nd Year. *Psychological Science*, *9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When Half a Word Is Enough: Infants Can Recognize Spoken Words Using Partial Phonetic Information. *Child Development*, *72*(4), 1003–1015. https://doi.org/10.1111/1467-8624.00331

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In Sekerina IA, Fernandez E, & Clahsen H (Eds.), Developmental Psycholinguistics: On-line methods in children's language processing (pp. 97–135). Amsterdam: J. Benjamins.

Ferreira, V. S. (2019). A Mechanistic Framework for Explaining Audience Design in Language Production. *Annual Review of Psychology*, *70*(1), 29–51. https://doi.org/10.1146/annurevpsych-122216-011653

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1276–1293. https://doi.org/10.1037/0096-1523.32.5.1276

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar Implicatures in Child Language: Give Children a Chance. *Language Learning and Development*, *8*(4), 365–394. https://doi.org/10.1080/15475441.2011.626386

Forbes, S., Dink, J., & Ferguson, B. (2021). *eyetrackingR*. R package version 0.2.0, http://www.eyetracking-r.com/

Fox, J. R. (1973). Encounter with anthropology. New York: Harcourt Brace Jovanovich.

Gatlin, B., & Wanzek, J. (2015). Relations Among Children's Use of Dialect and Literacy Skills: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *58*(4), 1306–1318. https://doi.org/10.1044/2015_JSLHR-L-14-0311

Goodman, K. S., & Buck, C. (1973). Dialect barriers to reading comprehension revisited. *The Reading Teacher, 27,* 6-12.

Green, L. J. (2002). African American English: a linguistic introduction. Cambridge University Press.

Green, L. J. (2010). Language and the African American child. Cambridge University Press.

Grice, P. 1989. Studies in the Way of Words. Cambridge, Mass.: Harvard University Press.

Grice, H.P. (1975). Logic and conversation. In *Speech Acts* (p.41-58). Brill.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55. https://doi.org/10.1016/j.cognition.2010.03.014

Haith, M. M., Wentworth, H., & Canfield, R. L. (1993). The formation of expectations in early infancy, *Advances in Infancy Research, 8,* 251-297.

Harber, J. R. (1977). Influence of presentation dialect and orthographic form on reading performance of Black, inner-city children. *Educational Research Quarterly, 2 (2), 9-16.*

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1179–1187. https://doi.org/10.1098/rstb.2010.0321

Harrington, J., Palethorpe, s., & Watson, C.I. (2007). Age-related changes in fundamental frequency and formations: A longitudinal study of four speakers. *Interspeech,* 2753-2756.

Hart, J. T., Guthrie, J. T., & Winfield, L. (1980). Black English phonology and learning to read. *Journal of Educational Psychology*, 72, 636–646.

Hartwell, P. (1980). Dialect interference in writing: A critical view. *Research in the Teaching of English, 14*(2), 101-118.

Hay, J., & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology*, *36*(1), 89–103. https://doi.org/10.1146/annurev.anthro.34.081804.120633

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3). https://doi.org/10.1515/TLR.2006.014

Heald, S. L. M., & Nusbaum, H. C. (2014). Talker variability in audio-visual speech perception. *Frontiers in psychology, 5,* 698.

Hoff-Ginsberg, E. (1994). Influences of mother and child on maternal talkativeness. *Discourse Processes,* 18, 105–117.

Holt, R. F., & Bent, T. (2017). Children's Use of Semantic Context in Perception of Foreign-Accented Speech. *Journal of Speech, Language, and Hearing Research*, *60*(1), 223–230.

https://doi.org/10.1044/2016_JSLHR-H-16-0014

Holt, Y., Méndez, L. I., Mills, M. T., & O'Brien, K. F. (2021). Phonological and Morphological Awareness Interventions with African American Children. *The Journal of Negro Education*, *90*(1), 13.

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739. https://doi.org/10.1037/a0016704

Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology*, *102*, 105–126. https://doi.org/10.1016/j.cogpsych.2018.01.004

Hurtado, N., Grüter, T., Marchman, V. A., & Fernald, A. (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Bilingualism: Language and Cognition*, *17*(1), 189–202. https://doi.org/10.1017/S136672891300014X

Hyter, Y. D., Rivers, K. O, & DeJarnette, G. (2015). Pragmatic language of African American children and adolescents: A systematic synthesis of the literature. *Topics in Language Disorders, 35*, 8-45.

Janssens, L., & Schaeken, W. (2012). The Role of Task Characteristics in Children's Scalar Implicature Production. *Proceedings of the Annual Meeting of the Cognitive Science Society, 34.*

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young Children Have a Specific, Highly Robust Bias to Trust Testimony. *Psychological Science*, *21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological*, *17*(9), 757–758.

Johnson, V.E. (2005). Comprehension of third person singular /s/ in AAE-speaking children. *Language, Speech, and Hearing Services in Schools, 36*(2), 116-124.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384. https://doi.org/10.1006/jpho.1999.0100

Kamide, Y., Altmann, G. T.M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133-156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kang, O., & Rubin, D. L. (2009). Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation. *Journal of Language and Social Psychology*, *28*(4), 441–456. https://doi.org/10.1177/0261927X09341950

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology, 104,* 105-114.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge, L., & Pascalis, O. (2005). Three-month-olds, but not newborn, prefer own-race faces. *Developmental Science*, *8*(6), pp. F31-F36.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy: Evidence of Perceptual Narrowing. *Psychological Science*, *18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Kibler, K., & Chapman, L. A. (2019). Six tips for using relevant texts in diverse classrooms. *The*

Reading Teacher, 72(6), 741-744.

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14,* 106-111.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580. https://doi.org/10.1073/pnas.0705345104

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent Trumps Race in Guiding Children's Social Preferences. *Social Cognition*, *27*(4), 623–634. https://doi.org/10.1521/soco.2009.27.4.623

Knoeferle, P., and Guerra, E. (2012). "What's non-linguistic visual context? A view from language comprehension," in What is a Context? Linguistic Approaches and Challenges, Vol. 196, eds R. Finkbeiner, J. Meibauer, and P. B. Schumacher (Amsterdam: John Benjamins Publishing), 9–32. doi: 10.1075/la.196.04mei

Kohler, C. T., Bahr, R. H., Silliman, E. R., Bryant, J. B., Apel, K., & Wilkinson, L. C. (2007). African American English Dialect and Performance on Nonword Spelling and Phonemic Awareness Tasks. *American Journal of Speech-Language Pathology*, *16*(2), 157–168. https://doi.org/10.1044/1058-0360(2007/020)

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2022). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications*, *7*(1), 7. https://doi.org/10.1186/s41235-022-00354-0

Labov, W. (2011). *Principles of linguistic change, volume 3: Cognitive and cultural factors.* John Wiley & Sons.

Labov, W. (2012). *Dialect diversity in America: The politics of language change.* University of Virginia Press.

Ladefoged, P., & Broadbent, D. E. (1957). Information Conveyed by Vowels. *The Journal of the Acoustical Society of America*, *29*(1), 98–104. https://doi.org/10.1121/1.1908694

Lanwermeyer, M., Henrich, K., Rocholl, M. J., Schnell, H. T., Werth, A., Herrgen, J., & Schmidt, J. E. (2016). Dialect Variation Influences the Phonological and Lexical-Semantic Word Processing in Sentences. Electrophysiological Evidence from a Cross-Dialectal Comprehension Study. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00739

Law, F., & Edwards, J. R. (2015). Effects of Vocabulary Size on Online Lexical Processing by Preschoolers. *Language Learning and Development*, *11*(4), 331–355. https://doi.org/10.1080/15475441.2014.961066

Law, F., Mahr, T., Schneeberg, A., & Edwards, J. (2017). Vocabulary size and auditory word recognition in preschool children. *Applied Psycholinguistics*, *38*(1), 89–125. https://doi.org/10.1017/S0142716416000126

Lawrence, D. (2017). The social perception of a sound change. *In the Proceedings of the 40th Annual Penn Linguistics Conference, 23,* 111-120.

Lee, O. E., & Braun, T. M. (2012). Permutation Tests for Random Effects in Linear Mixed Models. *Biometrics*, *68*(2), 486–493. https://doi.org/10.1111/j.1541-0420.2011.01675.x

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the Developmental Relationship Between Oral Language Skills and Reading Comprehension: It's Simple, But Complex.

Child Development, *89*(5), 1821–1838. https://doi.org/10.1111/cdev.12861

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Ma, D. S., Kantner, J., & Wittenbrink, B. (2020). Chicago face database: Multiracial expansion. *Behavior Research Methods*, *53*(3), 1289–1300. https://doi.org/10.3758/s13428-020-01482-5

Mahr, T., McMillan, B. T. M., Saffran, J. R., Weismer, S. E., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition, 142,* 345-350.

Mallinson, C. (2015). Sociolinguistics. In *Oxford Research Encyclopedia of Linguistics.*

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*(3), 367–380. https://doi.org/10.1016/j.cognition.2009.05.012

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177-190.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McGowan, K. B. (2015). Social Expectation Improves Speech Perception in Noise. *Language and Speech*, *58*(4), 502–521. https://doi.org/10.1177/0023830914565191

McMillan, B. T. M., & Saffran, J. R. (2016). Learning in Complex Environments: The Effects of Background Speech on Early Word Learning. *Child Development*, *87*(6), 1841–1855. https://doi.org/10.1111/cdev.12559

Melmed, P. J. (1973). Black English phonology: The question of reading interference. In J. L. Laffey & R. W. Shuy (Eds.), Language differences: Do they interfere? (pp. 70–85). Newark, DE: International Reading Association.

Metsala, J. L., Sparks, E., David, M., Conrad, N., & Deacon, S. H. (2021). What is the best way to characterise the contributions of oral language to reading comprehension: listening comprehension or individual oral language skills?. *Journal of Research in Reading*, *44*(3), 675-694.

Mitri, S. M., & Terry, N. P. (2014). Phonological awareness skills in young African American English speakers. *Reading and Writing*, *27*(3), 555–569. https://doi.org/10.1007/s11145-013-9458-z

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (2010). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Human Experimental Psychology, 50*(3), 528-559.

Munson, B. (2011). The influence of actual and imputed talker gender on fricative perception, revisited (L). *The Journal of the Acoustical Society of America*, *130*(5), 2631–2634. https://doi.org/10.1121/1.3641410

Munson, B., & Logerquist, M. (2017). Perceptual learning and gender normalization in fricative perception. *The Journal of the Acoustic Society of America, 141,* 3982-3982.

Münster, K., & Knoeferle, P. (2018). Extending Situated Language Comprehension (Accounts) with Speaker and Comprehender Characteristics: Toward Socially Situated Interpretation. *Frontiers in Psychology*, *8*, 2267. https://doi.org/10.3389/fpsyg.2017.02267

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: A preliminary investigation. *Journal of Child Language*, *25*(2), 343–365. https://doi.org/10.1017/S0305000998003444

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86,* 314-329.

Neys, W. D., & Schaeken, W. (n.d.). *WHEN PEOPLE ARE MORE LOGICAL UNDER COGNITIVE LOAD: DUAL TASK IMPACT ON SCALAR IMPLICATURE*. 14.

Norris, D. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210. https://doi.org/10.1016/S0093-934X(03)00053-1

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5,* 42-46.

Olson, R. H., Pomper, R., Potter, C. E., Hay, J. F., Saffran, J. R., Ellis Weismer, S., & Lew-Williams, C. (2020). Peyecoder: An open-source program for coding eye movements (Version v1.1.5). Zenodo. http://doi.org/10.5281/zenodo.3939234

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253–282. https://doi.org/10.1016/S0010-

0277(02)00179-8

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*(5), 1216–1226. https://doi.org/10.1037/0012-1649.43.5.1216

Piestrup, A. M. (1973).Black dialect interference and accommodations of reading instruction in first grade[Monograph No. 4]. University of California, Berkeley, Language and Behavior Research Laboratory.

Puranik, C., Branum-Martin, L., & Washington, J. A. (2020). The Relation Between Dialect Density and the Codevelopment of Writing and Reading in African American Children. *Child Development*, *91*(4). https://doi.org/10.1111/cdev.13318

Ratner, N. (2013). Why Talk with Children Matters: Clinical Implications of Infant- and Child-Directed Speech Research. *Seminars in Speech and Language*, *34*(04), 203–214. https://doi.org/10.1055/s-0033-1353449

Rips, L. J. (1975). Quantification and semantic memory. *Cognitive Psychology, 7*(3), 307-340.

Rowe, M. L., Coker, D., & Pan, B. A. (2004). A Comparison of Fathers' and Mothers' Talk to Toddlers in Low-income Families. *Social Development*, *13*(2), 278–291. https://doi.org/10.1111/j.1467-9507.2004.000267.x

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime reference guide. Pittsburge, PA: Psychology Software Tools.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465. https://doi.org/10.3758/s13428-017-0913-7

Seymour, H. N., & Ralabate, P. K. (1985). The acquisition of a phonologic feature of Black

English. *Journal of Communication Disorders*, 18, 139–148.

Seymour, H. N., Roeper, T. W., & de Villiers, J. (2003). *Diagnostic Evaluation of Language Variation–Screening Test*. The Psychological Corporation.

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15*(3), 436–447. https://doi.org/10.1111/j.1467-7687.2012.01135.x

Sharma, S. A., & Christ, T. (2017). Five steps toward successful culturally relevant text selection and integration. *The Reading Teacher, 71*(3), 295-307.

Shrodt, K., Fain, J. G., & Hasty, M. (2015). Exploring culturally relevant texts with kindergarteners and their families. *The Reading Teacher, 68*(8), 589-598. https://doi.org/10.1002/trtr.1363

Shutts, K., Banaji, M. R., & Spelke, E. S. (2009). Social categories guide young children's preferences for novel objects: Social categories guide preferences. *Developmental Science*, *13*(4), 599–610. https://doi.org/10.1111/j.1467-7687.2009.00913.x

Simons, H. D., & Johnson, K. R. (1974). Black English syntax and reading interference. *Research in the Teaching of English*, 8, 339–358.

Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, *30*(2), 191–205. https://doi.org/10.1016/0022-0965(80)90057-0

Speece, D.L., Roth, F.P., Copper, D.H., & De La Paz, S. (1999). The relevance of oral language literacy skills to early literacy: A multivariate analysis. *Applied Psycholinguistics, 20*(2), 16-190.

Stevenson, H. W., & Stewart, E. C. (1958). A developmental study of racial awareness in young children. *Child Development, 29*(3), 399-409. https://doi.org/10.2307/1126353.

Swingley, D., Pinto, J., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71,* 73-108.

Tavano, E., & Kaiser, E. (2010). Processing scalar implicature: What can individual differences tell us?. *University of Pennsylvania Working Papers in Linguistics*, *16*(1), 24.

Terry, J. M., Hendrick, R., Evangelou, E., & Smith, R. L. (n.d.). *34 Dialect Switching and Mathematical Reasoning Tests: Implications for Early Educational Achievement □*. 20.

Terry, J. M., Hendrick, R., Evangelou, E., & Smith, R. L. (2010). Variable dialect switching among African American children: Inferences about working memory. *Lingua*, *120*(10), 2463–2475. https://doi.org/10.1016/j.lingua.2010.04.013

Terry, J. M., Thomas, E. R., Jackson, S. C., & Hirotani, M. (2022). African American English speaking 2nd graders, verbal–s, and educational achievement: Event related potential and math study findings. *PLOS ONE*, *17*(10), e0273926. https://doi.org/10.1371/journal.pone.0273926

Terry, Nicole. (2008). Addressing African American English in Early Literacy Assessment and Instruction. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations*. 15. 54-61. 10.1044/cds15.2.54.

Terry, N. P. (2014). Dialect variation and phonological knowledge: Phonological representations and metalinguistic awareness among beginning readers who speak nonmainstream American English. *Applied Psycholinguistics*, *35*(1), 155–176. https://doi.org/10.1017/S0142716412000276

Terry, N. P., Connor, C. M., Johnson, L., Stuckey, A., & Tani, N. (2016). Dialect variation, dialect-shifting, and reading comprehension in second grade. *Reading and Writing*, *29*(2), 267–295. https://doi.org/10.1007/s11145-015-9593-9

Terry, N. P., Connor, C. M., Petscher, Y., & Conlin, C. R. (2012). Dialect Variation and Reading: Is Change in Nonmainstream American English Use Related to Reading Achievement in First and Second Grades? *Journal of Speech, Language, and Hearing Research*, *55*(1), 55–69. https://doi.org/10.1044/1092-4388(2011/09-0257)

Terry, N. P., Connor, C. M., Thomas-Tate, S., & Love, M. (2010). Examining Relationships Among Dialect Variation, Literacy Skills, and School Context in First Grade. *Journal of Speech, Language, and Hearing Research*, *53*(1), 126–145. https://doi.org/10.1044/1092-4388(2009/08-0058)

Thomas-Tate, S., Washington, J., & Edwards, J. (2004). Standardized Assessment of Phonological Awareness Skills in Low-Income African American First Graders. *American Journal of Speech-Language Pathology*, *13*(2), 182–190. https://doi.org/10.1044/1058-0360(2004/018)

Treiman, R., & Kessler, B. (2004). The case of case: Children's knowledge and use of upper- and lowercase letters. *Applied Psycholinguistics*, *25*(3), 413–428. https://doi.org/10.1017/S0142716404001195

Troutman, D. E., & Falk, J. S. (1982). Speaking Black English and Reading—Is There a Problem of Interference? *The Journal of Negro Education*, *51*(2), 123. https://doi.org/10.2307/2294677

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Reading Assessment.

U.S. Department of Education. Institute of Education Sciences, National Center for Education

Statistics, National Assessment of Educational Progress (NAEP), 2022 Reading Assessment.

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*(1), 340–350. https://doi.org/10.1037/a0032192

Venker, C. E., Pomper, R., Mahr, T., Edwards, J., Saffran, J., & Ellis Weismer, S. (2020). Comparing Automatic Eye Tracking and Manual Gaze Coding Methods in Young Children with Autism Spectrum Disorder. *Autism Research*, *13*(2), 271–283. https://doi.org/10.1002/aur.2225

Voeten, C.C. (2022). *permutes,* R package version 2.6*.* https://cran.r-project.org/package=permutes

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00546

Washington, J. A., Branum-Martin, L., Sun, C., & Lee-James, R. (2018a). The Impact of Dialect Density on the Growth of Language and Reading in African American Children. *Language, Speech, and Hearing Services in Schools*, *49*(2), 232–247. https://doi.org/10.1044/2018_LSHSS-17-0063

Washington, J. A., Branum-Martin, L., Sun, C., & Lee-James, R. (2018b). The Impact of Dialect Density on the Growth of Language and Reading in African American Children. *Language, Speech, and Hearing Services in Schools*, *49*(2), 232–247. https://doi.org/10.1044/2018_LSHSS-17-0063

Washington, J. A., & Craig, H. K. (1998). Socioeconomic Status and Gender Influences on

Children's Dialectal Variations. *Journal of Speech, Language, and Hearing Research*, *41*(3), 618–626. https://doi.org/10.1044/jslhr.4103.618

Weatherhead, D., Kandhadai, P., Hall, G.D., & Werker, F. (2021). Putting mutual exclusivity in context: Speaker race influences monolingual and bilingual infants' word-learning assumptions. *Child Development, 92*(5), 1735-1751.

Weatherhead, D., & White, K. S. (2018). And then I saw her race: Race-based expectations affect infants' word processing. *Cognition*, *177*, 87-97.

Weatherhead, D., & White, K. S. (2020). Toddlers link social and speech variation during word learning. *Developmental Psychology*, *57*(8), 1195–1209. https://doi.org/10.1037/dev0001032

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(Issue 11, Supplement 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian Infants Scan Own- and Other-Race Faces Differently. *PLoS ONE*, *6*(4), e18621. https://doi.org/10.1371/journal.pone.0018621

Williams, K. T. (2006). Expressive Vocabulary Test (2nd ed.). Minneapolis, MN: Pearson Assessments.

Winn, M. B., Rhone, A. E., Chatterjee, M., & Idsardi, W. J. (2013). The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00824

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The Relationship Among Receptive and Expressive Vocabulary, Listening Comprehension, Pre-Reading Skills, Word Identification Skills, and Reading Comprehension by Children With Reading Disabilities. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1093–1109. https://doi.org/10.1044/1092-4388(2007/076)

Wolfram, W. (1971). Black and White speech differences revisited. In W. Wolfram & N. H. Clarke (Eds.), Black-White speech relationships (pp. 139–161). Washington, DC: Center for Applied Linguistics.

Wolfram, W. (2001) Reconsidering the sociolinguistic agenda for African American English. In S.L. Lanehart (ed.) Sociocultural and Historical Contexts of African American English (pp. 321–62). Amsterdam/Philadelphia: Benjamins.

Wolfram, W., & Schilling, N. (2015). American English: Dialects and variation (3rd ed.). Hoboken, NJ: Wiley-Blackwell.

Xiao, M., Wang, R., & Chan-Olmsted, S. (2018). Factors affecting YouTube influencer marketing credibility: A heuristic-systematic model. *Journal of Media Business Studies*, *15*(3), 188–213. https://doi.org/10.1080/16522354.2018.1501146

Yang, X., & Krajbich, I. (2020). *Webcam-based online eye-tracking for behavioral research* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/qhme6

Yoon, E., Charles Wu, Y., & Frank, M. (2015). Children's online processing of ad-hoc implicatures. *Cognitive Science.*

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L.A., Voinescu, L., & Collins, M. (1996). “Wide-eyed” and “crooked-faced”: Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin, 22*(12), 1-12.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832