Children's Learned Associations with Voice: Perspectives on Children's Speech Perception in Language Acquisition

Nicholas P. Moores, Kevin B. McGowan, Meghan Sumner,
& Michael C. Frank
Departments of Linguistics and Psychology, Stanford University

1 Abstract

Speech perception provides an avenue with which to study children's burgeoning ability to comprehend social information in the speech stream. The speech signal contains not only phonemic information necessary for word recognition but abundant "indexical information" about the talker, including gender (Perry, Ohde, & Ashmead, 2001? Goldinger, 1998; Johnson et al., 1999; Johnson, 2006). A growing body of work has begun to investigate the effects of listeners' associations between talkers' social identity and the phonetic cues in talkers' voices, demonstrating that by adulthood listeners use talker information to make social inferences about a talker's likely behavior (Van Berkum et al., 2008), especially when they expect talker identity to be useful or find it to be a reliable cue (Creel, Aslin, & Tanenhaus, 2008). Recent work suggests that children use acoustic cues to talker identity to constrain comprehension of spoken language (Creel, 2012), though the way in which children learn to integrate social knowledge with information from talker voice remains poorly understood. In this paper, we test the hypothesis that children are able to disambiguate between objects with gendered associations (a men's pair of gloves and a women's pair of gloves, say) based on talker voice. We explore children's use of talker indexical information to infer speaker meaning through an experiment in which children interact with a web page on an iPad. Over 24 trials, children ages 3-5 were shown series of four images and asked by talkers to find one of the objects by clicking on it. During half of the trials, children heard a male talker's voice, and during the other half they heard a female talker's voice. In addition, half of the trials were non-competitor trials in which there was only one image of the talker's referent (hearing a man's voice and seeing a man's glove, say). The other half were competitor trials in which the target image competed with a variant which would stereotypically belong to a speaker of the opposite gender (hearing a man's voice but choosing between both a man's glove and a woman's glove). Children's reaction time between the utterance of the target word and their click on an image was logged, as was their choice of image. Preliminary data show that by the age of 5, children regularly integrate phonetically-cued socially indexical talker information with their social knowledge of speaker characteristics to guide their interpretations of speaker meaning in a real world paradigm. From ages 3 to 5, the ability to disambiguate increases significantly. All ages are extremely good at the task in non-competitor trials. Interestingly, in competitor trials, as age increases, children are faster and more accurate in disambiguating the voice cues. These results suggest not only that children make use of socially-nuanced talker-specific acoustic information by a young age, but reveal a robust understanding of gender stereotypes that guide their daily interactions with interlocutors and may bear on their linguistic and social development.

2 Introduction

I propose to examine the effect of gender-specific acoustic information on voice on the inferences children make in conversation. Previous research has demonstrated that children remain sensitive to much phonetic detail and acoustic information previously thought to be generalized, and that they use this fine-grained information to construct social categories and to access lexical candidates that would seem appropriate given the speaker's social identity. Preliminary data with adults has also demonstrated that in online processing of the speech stream adults use the gender- and age-specific acoustic characteristics retrieved from the speaker's voice to attend to objects that are semantically connected to the speaker given the speaker's age and gender. This effect has not yet been researched in children, however. The goal of my proposed study therefore is to replicate these findings in a sample in which the effects of talker-specific voice characteristics on attention to socially-graded objects has not yet been examined, young children. In this study I will examine whether children, who are themselves beginning to construct the social categories they will carry with them into adulthood, are yet able to use age- and gender-specific voice information to guide their attention in everyday life.

I hypothesize that by the age of five, children will be able to, based upon the vocal cues they hear (whether it's a man's, woman's, or child's voice), attend to the critical object, that which is appropriate given the speaker's age and gender. I further hypothesize that they will be slower to attend to the correct object when there is a competitor for age than when there is a competitor for gender, and that children will attend most quickly to the target object when there is no competitor at all.

3 Experiment 1

3.1 Method

Participants. Participants will be children age 3-5 who attend Bing Nursery School.

Stimuli. Our visual stimuli will be normed through extensive survey on Amazon Mechanical Turk, where raters will be asked, given only the picture the child subject will see in the actual experiment, to name the object, and to rate in percentage how strongly they would agree that the item belongs to a man, belongs to a woman, and belongs to a child. This step will help ensure that the images used do not introduce unexpected confounds into the experimental design. Thus by using adults as a kind of control, we can see how well adults would be able to understand that a given item is the item we are naming it as, and that that item clearly belongs to either a man, woman, or child.

Equipment. This project will utilize the web programming experimentation techniques used in the Language and Cognition Lab, wherein the experiment is crafted using HTML, CSS, JavaScript (with jQuery), and PHP, and is run on an internet-connected iPad in Guided Access mode (the subject cannot launch another iPad program or exit the webpage on which the experiment page is loaded, displaying a visual-world paradigm. The experiment webpage on the iPad receives the conditions for that subject from the lab server and submits the results to a spreadsheet on the lab server.

Procedure. When the experiment begins, five colored dots will appear on the iPad screen; the dots turn to x's when the subject clicks on them. The child subject must successfully click on all of them in order to proceed to the actual experiment; this ensures that the child understands and is familiar with clicking an iPad screen as a mode of interaction with the experiment.

Once the subject has clicked on all five dots, twelve trials ensue in random order, in which a voice plays asking for the subject to find and click on an object they're looking for. The voice will either belong to an adult male, an adult female, or a female child. Four images appear in each trial. Each trial will feature a different item that the speaker is looking for, and throughout the experiment the child will only hear one person's voice. During half of the trials, the item the speaker is looking for will appear with three images of other distractor items, such that if the speaker is asking for a *bag*, there will only be one *bag* displayed on the screen (the target picture the child should pick) with three other items. During all other trials, two items matching the speaker's description will appear with two images of other distractor items, such that if the speaker is asking for a *bag*, there will be two *bags* displayed on the screen with two other distractor items. These are the competitor or critical trials. Depending on the subject, the target item and the competitor item will be different in that (according to our image norming data) they would typically be owned by people of the same age but different genders, or they would typically be owned by people of the same gender but different ages. We seek to find by what age children can successfully use the voice of the speaker in these critical trials to pick the item that would be owned by someone of both the speaker's age and gender. The experimenter selects a set of 12 trials to use, but the order of the trials is conducted randomly.

3.2 Results

3.3 Discussion

4 Experiment 2

Coming soon!

5 General Discussion

6 Broad Arguments

Given that children use voice cues to talker identity to constrain comprehension of spoken language (Creel 2012), and that the speech signal contains abundant "indexical information" about the talker, including gender (Perry, Ohde, & Ashmead, 2001). However, children have more trouble than adults in integrating cues (Zelazo, Frye, & Rapus, 1996).

Could talk some about acoustic-encoding account of talker-specificity effects and semantic-encoding account of talker-specificity effects. If we follow the acoustic-encoding account, these sounds I am hearing are a more precise acoustic match between the current speech event and the previous encoded speech event. If we follow the semantic-encoding account, these sounds I am hearing have talker-specific advantages in processing that stem from activation of knowledge about people, and these voice characteristics are one way to activate this knowledge (Creel & Tumlin, 2011). This would be in line with an exemplar-based model of resonance (Johnson 2006). In fact, in a series of eye-tracking experiments, Creel (2012) demonstrated that children use voice characteristics to access their knowledge about people – thus augmenting their comprehension. Furthermore, Creel (2012) showed that children can use talker-specificity to learn preferences related to the individual even when that individual's preferences aren't buttressed by gender stereotypes. ***put in more support here*** this all supports the semantic-encoding account of talker-specificity effects.

Prior work has shown that adults use talker information to make inferences about a talker's likely behavior (Van Berkum et al. 2008) but little work has yet been done to investigate the extent to which children make these same inferences. If we follow a semantic encoding account of talker-specificity effects, per the evidence given above, we can presume that children eventually do the same. Little work has been done to show when children begin to use this information as well. We here demonstrate a developmental trajectory for children's use of semantic encoding of talker-specificity, with 3-year-olds operating slightly below chance, 4-year-olds operating slightly above chance, and 5-year-olds operating significantly above chance but not yet at a level of adult proficiency.

FROM VAN BERKUM ET AL 2008: *Event-related brain responses revealed that the speaker?s identity is taken into account as early as 200-300 msec after the beginning of a spoken word, and is processed by the same early interpretation mechanism that constructs sentence meaning based on just the words.*

This finding is difficult to reconcile with standard ?Gricean? models of sentence interpretation in which comprehends initially compute a local, context-independent meaning for the sentence (?semantics?) before working out what it really means given the wider communicative context and the particular speaker (?pragmatics?).

Because the observed brain response (N400s) hinges on voice-based and usually stereotype-dependent inferences about the speaker, it also shows that listeners rapidly classify speakers on the basis of their voices and bring the associated social stereotypes to bear on what is being said. According to our event-related potential results, language comprehension takes very rapid account of the social context, and the construction of meaning based on language alone cannot be separated from the social *aspects of language use. The linguistic brain relates the message to the speaker immediately.*

In psycholinguistics, assuming a Gricean standpoint has led to an analysis that has evolved into the standard two-step model of language interpretation. In this model, listeners (and readers) first compute a local, context-independent meaning for the sentence, and only then work out what it really means given the wider communicative context and the particular speaker (Lattner & Friederici, 2003; Cutler & Clifton, 1999; Sperber & Wilson, 1995; Fodor, 1983; Grice, 1975). Mismatches between message and speaker would be detected in the second step only, in slow pragmatic computations that are different from the rapid semantic computations in which word meanings are combined (e.g., Lattner & Friederici, 2003). From a design perspective, however, it would make sense for the linguistic brain to take the speaker into account right from the start. After all, human language has evolved to support social, interpersonal interaction.

Also, recent linguistic research suggests that the computation of a context-free sentence meaning is, in fact, highly problematic, and that linguistic meaning is always colored by the pragmatics of the communicative exchange (Kempson, 2001; Perry, 1997; Clark, 1996). The meaning of so-called indexicals such as ?I? and ?you,? for example, inevitably depends on the communicative situation (e.g., Perry, 1997), and upon closer analysis, so does the meaning of apparently self-sufficient words such as ?garage? or ?Kensington Gardens? (Kempson, 2001; Clark, 1996).

These analyses are at odds with the standard two-step model of interpretation. Instead, they suggest a one-step model in which knowledge about the speaker is brought to bear immediately by the same fast-acting brain system that combines the meanings of individual words into a larger whole.

Hanna and Tanenhaus (2004) showed that listeners who were asked to hand over something to the speaker were rapidly sensitive to whether the latter could have picked up the item himself or herself. Such eye tracking findings show that, in conversation, listeners rapidly relate the message to characteristics of the speaker.

The critical prediction of this model is that because semantic information provided by the words in a sentence and voice- conveyed information about the identity of the speaker are handled by the same early sense-making process, semantic anomalies and speaker inconsistencies will generate the same type of ERP effect, an N400 effect, doing so in the typical latency range for N400 amplitude modulations. The two-step model of semantic interpretation makes a different prediction: If contextual information about the speaker is handled in a dis- tinct second phase of interpretation, then speaker inconsistencies should elicit a delayed and possibly quite different ERP effect. According to our ERP results, the brain integrates message and speaker *very rapidly, within some 200-300 msec after the acoustic onset of a relevant word. Also, speaker inconsistencies elicited the same type of brain response as semantic anomalies, an N400 effect. That is, voice-inferred information about the speaker is taken into account by the same early language interpretation mechanisms that construct "sentence-internal? meaning based on just the words. Our findings therefore demonstrate that, as far as the brain is concerned, linguistic meaning depends on the pragmatics of the communicative situation right from the start. However, by revealing an equally immediate impact of what listeners infer about the speaker, the present results add a distinctly social dimension to the mechanisms of on-line language interpretation. Language users very rapidly model the speaker to help determine what is being said. This makes sense, as language evolved in face-to-face social interaction and, importantly, requires close coordination among interlocutors (Clark, 1996).*

As far as all this goes, this pretty successfully argues that the social knowledge about a speaker is immediately activated; it doesn't come in a separate, second pragmatic parsing after an immediate semantic context-free parsing. Van Berkum is showing us that by adulthood, we immediately take into account social inferences based on stereotypes to parse speaker meaning. Creel is telling us, importantly, the role of the development of children's use of talker information from the speech stream. But how soon are they able to use their knowledge of social categories and stereotypes to infer speaker meaning? Our work suggests that by the age of 5, children have no problem using their social knowledge of speaker characteristics (based on social categories and stereotypes that have been constructed for them or that they have constructed) to make these inferences in everyday conversation. If talker information shaped by social knowledge is present in the immediate parsing of the speech stream, understanding how this ability develops would shed new light onto the social aspects and connotations of children's language acquisition.
Our interpretation is supported by exemplar models of the lexicon (Goldinger, 1996; Johnson, 2006). Johnson 2006 also demonstrates how speakers perform social identity in a substantially linguistic way.

However, rather than rapidly disregarding voice-based cues to who the speaker is, listeners, in fact, use these cues in the earliest stages of speech signal processing (for reviews, see Johnson, 2005; Nygaard, 2005; Thomas, 2002). For example, listeners perceive vowels differently depending on whether a preceding stretch of filtered speech is shifted up or

down in frequency to suggest a male or female speaker (Remez, Rubin, Nygaard, & Howell, 1987), and the perception of an ambiguous syllable-initial fricative is systematically modulated by whether the rest of the syllable is spoken by a man or a woman (Strand, 1999).

These speech perception findings demonstrate that listeners can extract and use information about the speaker from the acoustic signal very rapidly. Furthermore, merely showing the face of a male or female speaker also affects fricative and vowel perception (Johnson, Strand, & D?imperio, 1999; Strand, 1999). This suggests that speaker identity is taken into account in the earliest stages of speech perception in a way that goes beyond a simple within- modality mechanism.

FROM JOHNSON ET AL 1999: *Talker normalization theory assumes that listeners? subjective, abstract impressions of talkers play a role in speech perception. The gender of a visually presented face affects the location of the phoneme boundary between two vowels in the perceptual identification of a continuum of auditory-visual stimuli ranging from hood to hud. This effect was found for both ?stereotypical? and ?non-stereotypical? male and female voices.*

The results from these experiments suggest that listeners integrate abstract gender information with phonetic information in speech perception. This conclusion supports the talker normalization theory of perceptual speaker normalization.

FROM CREEL 2010: *No two people sound alike. Some research indicates that this poses a challenge for language processing (Mullennix, Pisoni, & Martin, 1989; Nusbaum & Morin, 1992). However, it may also provide additional, helpful information to the comprehender. That is, knowing who is talking can provide useful information in processing spoken language. For instance, adult listeners make different predictions about upcoming information in a sentence depending on who is speaking it (Van Berkum, et al 2008), suggesting they have particular semantic associations with certain voice characteristics (e.g., a child?s voice vs. an adult?s voice). Thus, acoustic differences among talkers potentially have rich semantic associations (Geiselman & Crawley, 1983). But how long does it take the developing language learner to form and use these associations in comprehending language?... The current work suggests that preschool-aged children are similarly able to use talker acoustics to calculate listeners' likely and unlikely referents.*

The current work, as well as Van Berkum's, fits nicely with a perspective on *language processing (Kamide, Altmann, & Haywood, 2003; Bicknell et al 2008) in which comprehends use any available linguistic and nonlinguistic cues to construct event representations on-line. Acoustic information linked to talker identity is apparently useful in constructing event representations, and is a robust enough cue that preschool-aged children can use it rapidly on-line (Snedeker & Trueswell, 2004). Creel talks about using voice information of talker gender to guess whether the speaker prefers pink or blue as a low-level auditory-visual cue correspondence rather than knowledge of an individuals, though they found instead that children were actually accessing talker identity of the agent speaking (Experiment 2)* **Our study kind of answers this question as well, but in the sense that we're not just talking about individual talker identity, but inferences about whole groups of people based on perceptions of social categories**

FROM CREEL & TUMLIN 2011: *Creel was interested in both listeners? use of acoustic-match information, where word forms are stored in acute acoustic detail, and their use of talker-semantic information ? linkages between talkers and the things they talked about. The studies found evidence for both types of information.*

FROM CREEL 2010: *Children seem adept at processing prosodic information. Snedeker and Yuan (2008) showed that children were sensitive to a speaker's intonational phrase boundaries in their interpretations of prepositions-phrase attachment. Ito, Jincho, Minai, Yamane, and Mazuka (2009) and Bibyk, Ito, Wagner, and Speer (2009) found that children as young as 6 years use pitch accent to constrain upcoming referents to a set of items contrasting on the pitch-accented dimension. These studies suggest that children attend to non-phonemic sound patterns that cue differences in meaning.*

Children seem to have more difficulty processing cues to vocal affect. Morton and Trehub (2001) found that when vocal affect conflicts with verbal content, children cannot ignore the verbal content when reporting the talker?s affect (reporting the first sentence as sounding happy even though its intonation was sad, and the second as sounding sad even though the intonation was happy). Nonetheless, recent work by Berman, Graham, and Chambers (2009) using eye tracking, a more sensitive, implicit measure, suggests that children associate positive and negative vocal affect cues with positively- and negatively-valenced pictures (e.g. intact vs. broken dolls).

Children may be using these cues to make associations between sound properties and semantic attributes. For instance, pitch accent seems to semantically activate contrast sets. In the vocal emotion case, children

might have associations between sad vocal cues and non-intact objects (Berman et al., 2009). This leaves open whether children are able to use non-phonemic acoustic information in the speech signal to make high-level inferences about the perspective taker.

In sum, children show some ability to glean semantic information from two non-phonemic acoustic information sources, prosody and vocal affect. Thus, one might expect that children would gain semantic information from non-phonemic acoustic cues to talker as well. However, it is not clear that children can go so far as to use it to invoke a particular talker?s perspective.

FROM CREEL, ASLIN, & TANENHAUS 2008: *Different talkers exhibit different fundamental frequencies (Van Lancker, Kreiman, & Wickens, 1985), speaking rates (Van Lancker et al., 1985), voice onset times (Allen, Miller, & DeSteno, 2002), etc. However, none of these factors influences word meaning. Phonemic categories, then, are key to identifying lexical entries, and other attributes of the speech signal are viewed as irrelevant and even detrimental to the process of spoken word recognition. This traditional perspective deals effectively with the first of two difficult problems in spoken word recognition: limited storage capacity. By storing only phonemic information, a large number of lexical entries can be represented with a small, compact set of symbols.*

Creel Tumlin 2011 showed that listeners may store acoustic-match information with little effort, but store talker-semantic information when it is attended or is highly diagnostic of a response, and perhaps when memory demands are at a minimum.

The study demonstrates that listeners encode (or utilize) what seems to be talker-semantic information very differently depending on the task they are performing.

When the listener expects talker identity to be useful, either because it is necessary to succeed in the task (Experiment 3) or because it is an extremely-reliable response cue (Experiment 4), the listener will use talker-semantic information in on-line processing. It should be kept in mind that the 'task utility' of talker-semantic information in the real world may change from situation to situation, and may be greater than that demonstrated in our study.

It is convenient to dichotomize variability in the speech signal into rapidly-varying phonemic characteristics (voice onset time, formant transitions) vs. slow-varying talker-related characteristics (f0, formant frequency range), which would neatly lateralize speech leftward and talker rightward. Extending this to our proposed acoustic/semantic distinction, identification of talkers might proceed from coarse-grained analyses, and identification of words might proceed from fine-grained analyses.

Complicating this tidy picture, talker variability cross-cuts both coarse and fine acoustic time scales, and listeners can use both coarse and fine-grained information to identify talkers (Allen et al., 2003; Fellowes et al., 1997; Remez et al., 1997). This suggests that talker representations and word representations, even if they function separately, must share some information. This might mean that talker representations, while including some fine-grained information, are biased toward coarse-grained information more strongly than fine-grained information, with the reverse being true for speech processing.

FROM CREEL 2012: *Human listeners are able to use acoustic cues to talker to constrain on-line language processing as early as the preschool years. Preschoolers appear to accomplish this by using voice characteristics to access representations of talkers? (or agents?) mental states, specifically, their color preferences. Preschoolers rapidly learn and use simple preference knowledge about new talkers. However, neither preschoolers nor adults take advantage of vocal cues to identify two newly learned characters matched for gender, age, and dialect. For adults, this likely results from inattention rather than inability to discriminate the two voices. In children?s case, difficulties are likely due to failure to discriminate, which itself may be due to inattention to talker characteristics.*
This perspective is consistent with an account by which listeners learn what elements of the speech signal they should direct attention toward in different contexts. This includes exemplar-style theories of sound representation in language: that listeners store acoustic attributes of speech whole (Goldinger, 1998), including both phonemic variability and talker variability. What occurs over development, rather than a progressive narrowing of information intake, is the emergence in the exemplar cloud of phoneme-related and talker-related regularities via perceptual learning.

Categorical perception has been defined as the ability to discriminate between-category contrasts better than within-category contrasts (Liberman 1970). By at least 7 months of age, infants do appear able to categorize male and female voice tokens (Miller 1983). Children at least notice gender cues by 3 to 5 months of age, but cannot match dynamic faces and voices based on gender cues until the second half of their first year of life (Patterson & Werker, 2002).

Meghan's Two Cents: we're asking at what point do you use accumulated social knowledge? Highlight the benefits of Creel's study, showing this is learnable in the short term and it's cleaner than what we are doing. We can't say as well as she can what type of learning led to this but we can say co present voices with objects. That's the hardest part – to give her credit, say from that we know that but we know from birth we're getting an input telling you about sounds and words but also telling you about talkers. We're talking about social stereotypes coming from voice cues to make those inferences. Creel can't tell you anything about social interpretations – she can tell you talker a likes white and talker b likes black but she can't say anything about social representations at all. Her's is more controlled and ours is more impactful. It's not a voice base but it's a gender stereotyped base. Write this up as this is more interesting. Do we have any evidence at all that kids are able to take patterns and infer talker meaning reliably? This is spoken language understanding in that they're understanding inference by the talker.

A cost of messiness but there's a beauty of ecological validity – what links do you already have between voice cues and social cues and how do they guide your interpretations.

So we can say that our study at one level is looking at the fact that children use talker information in voice to access stereotypical information from social categories from a very early age (speech perception), and that at another level is looking at how our results bear on the discussion of how children are actually constructing gender stereotypes and social categories more broadly (socialization).

For our speech perception section, talk about how we're building on and responding to Creel, van Berkum, Johnson, Sumner, Goldinger. For our socialization dialogue, talk about how we're building on and responding to Andersen, Gleason, Greif

7 Lit Review: Talker Identity in Child and Adult Speech Perception

Children are quite sensitive to the acoustic signal, and use it well to learn to speak and understand what will become their native language. Prior research, however, has demonstrated that children aren't excellent at recognizing voices. Voices are somewhat like faces in their ability to convey both categorical information and individual identity. The speech signal contains indexical information about the talker including gender, age, femininity, sexual orientation, region of origin, socioeconomic status, and individual identity. The knowledge of any of these would help you figure out what the person is going to say, or at least constrain your guesses of what they are going to say. Children are sensitive to acoustic variability in word-recognition and word-production tasks, suggesting that they may be sensitive to talker variability in comprehension as well, and studies with adults have shown that talker-specificity aids in recognition of target words, suggesting that these effects result from precise acoustic encoding of words that contains both phonemic information and voice characteristics. Adults have been shown to make inferences based on talker identity during language processing, and there has been shown to be semantic mismatch when adults hear a child say they would like 'a glass of wine' than when an adult says it. This suggests that language comprehenders, at least by adulthood, encode talker information *semantically*, which seems at odds with the precise acoustic encoding account. But do children know how to do this?

In their first year, children have very acoustically specific representations of word forms: 7.5 month olds have difficulty generalizing a newly familiarized word over gross changes in talker (Houston & Jusczyk, 2000), vocal emotion (Singh, White & Morgan, 2008), and accent (Schmale & Seidl, 2009). This early sensitivity to acoustic form in the first year of life is consistent with an account of phonological development where young learners move from language-general to language-specific sound sensitivity, gradually focusing attention on information relevant to meaning in the target language (Werker & Tees, 1984). But can children use acoustic cues to identify individuals? Identifying an individual via an acoustic cue would involve both perceiving said cues and mapping them onto particular individuals or groups. Such mappings have been suggested; 5-year-olds state preferences to be friends with people who speak an acoustically familiar language (Kinzler, Dupoux, & Spelke, 2007). It has also been shown that children associate an unfamiliar language with unfamiliar-looking clothing, dwellings, and people, which suggests an ability in children to associate acoustic information with semantic associates.

Children can enhance their understanding of spoken language by knowing what person, or what sort of person, is talking to them, but earlier work calls into question how well children can use voice to access this information. Five to six year old children often process cues more slowly than adults, or fail to integrate cues with linguistic information. Additionally children use paralinguistic information – prosodic patterns and vocal-emotional information – less adeptly than adults. This suggests that preschool children's associations of voice with meaning may be present but somewhat fragile. This work aims to find how well in fact children can use semantic associations with voice to make pragmatic inferences which guide their language comprehension and acquisition, especially when those associations are dependent on talker-specific acoustic information. Knowing if children do indeed use learned semantic associations with voice to make pragmatic inferences during language acquisition will greatly inform further work seeking to understand the interaction of children's developing linguistic skills and their social cognition.

Notes on Gender Stereotypes.

Gleason (1973). By age 3 or 4, children are aware of at least some aspects of these relationships about who you ask what about, and how you ask it to that person depending on your relative rank with other speakers.

Gleason (1975). fathers tend to use more imperatives when speaking to sons and daughters and suggested that, early in life, boys, unlike girls, become used to giving and receiving orders. Supports Bellinger & Gleason 1982 Hypothesis 2 that parental modeling is the important process involved in children?s acquisition of sex-associated ways to request action. ALSO, shows that imperatives account for a greater percentage of males? than females? utterances to young children across a variety of settings. Supports Bellinger & Gleason 1982 Hypothesis 1 that parents use different forms to boys and girls, who, in turn, learn to speak as they have been spoken to.

Greif (1980). Parents interrupt preschool girls more frequently than they interrupt preschool boys. Supports Bellinger & Gleason 1982 Hypothesis 3 that the directive forms that parents address to young girls and boys reflect the ultimate status that males and females have in our society.

Bellinger & Gleason (1982). Broader phenomenon where fathers not only produce more imperatives than mothers, but more directive speech acts in general. We found that mothers and fathers select differently from among the forms which can be used to express directive intent. Their question revolved mostly around socialization in terms of requesting action (they only saw evidence for Hypothesis 1 there), but evaluating the triune forces of parental modeling, the differences in CDS between mothers and fathers, and along with that the ultimate status that males and females have in society (marked v. unmarked, among other things) remain important and interesting even when generalized to broad-spectrum socialization of children, a side effect of which is the acquisition of gender stereotypes.

McGillicuddy-De Lisi (2002). Has even more gender stereotypes work and a variety of other perspectives on gender behavior. Sachs (1987) observed same-sex 2 to 5 year old child dyads in pretend play as doctors; boys almost chose the role of doctor for themselves and assigned their partners to other roles, whereas girls usually asked their partners what role they would like to play. Regardless of role played, however, boys used more imperatives and prohibitions than girls, and girls made more use of tag questions and joint utterances. McGillicuddy-De Lisi (2002): Most of the stereotypes that we have about differences between men and women are relatively accurate (Swim, 1994). Language and other cognitive skills are developed through human interactions and these interactions include emotional content communicated in a social setting. Thus, a full understanding of cognitive sex differences will also need to include the development of emotional understanding, which may be the first language to develop. [Swim, J.K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. Journal of Personality and Social Psychology, 66, 21-36.]

An in-group/out-group social psychology that likely evolved in the context of competition between relatively small kin-based groups more likely than not provided the foundation for the evolution of social ideologies (Alexander, 1990). These ideologies are particularly important, because they appear to be the basis for the formation of large nation-states, that is, the social organization of individuals who have never met, and never will, and thus are unable to develop one-on-one personal relationships (Geary, 1998). Such ideologies define the mutual self-interest of individuals who comprise groups that are larger than functional villages in preindustrial societies and are the basis for large-scale between-group conflict. In fact, the competitive advantage associated with group size was the likely pressure, after the emergence of language and shared belief systems (e.g., origin myths), that resulted in the evolution of the tendency of humans to form and rally around such ideologies (Alexander, 1990). In support of this view is the finding that people show an enhanced endorsement of in-group ideologies and harsher evaluations of out-group members under conditions that imply a threat to one?s mortality (Arndt, Greenberg, Pyszczynski, & Solomon, 1997). As predicted in the book?s treatment of development, many of the sociocognitive sex differences are evident early in life, are manifested in social and play patterns, and some of these have been linked to prenatal exposure to sex hormones.
Andersen (1990). Children learning to use different registers as appropriate in different social situations as they become socialized. Give landmarks from notes. Look to her references for other sources. Further evidence that children have acquired knowledge and mastery of gender-specific language can be found in young children's ability to role play adult males and females. As noted earlier, when children between the ages of 4 and 7 were asked to enact parents' speech, their pitch was significantly deeper as fathers than as mothers. In addition, when they spoke for the father puppet they modified their language across domains. They altered their phonology as fathers by using deeper voices, were louder (sometimes yelling), and used back and lower vowels to produce an accent that could be thought of as sinister. When they pretended to be mothers they used higher pitch, exaggerated intonation, and chose stereotypically female vocabulary. In their discourse, pretend fathers spoke about work, business meetings, how tired they were from working at the computer, and

how they had to fire their secretary. In the role of mother, they complained about being exhausted from their errands.

Philips, Steele, & Tanz (1987). Has more Gleason and gender stereotypes work.

Notes on Gender Stereotypes and Social Cognition.

Bargh, Chen, Burrows (1996): Prior research shows that trait concepts and stereotypes become active automatically in the presence of relevant behavior or stereotyped-group features. These activated stereotypes have effects on participants such that they behave according to the stereotype following the experiment in which the stereotype was activated.

Greenwald, A. G., & Banaji, M. R. (1995): A stereotype is a socially shared set of beliefs about traits that are characteristic of members of a social category. Whereas an attitude implies a consistent evaluative response to its object, a stereotype may encompass beliefs with widely diverging evaluative implications. For example, the stereotype of members of a certain group (cheerleaders, say) may simultaneously include the traits of being physically attractive (positive) and unintelligent (negative). Stereotypes guide judgement and action to the extent that a person acts toward another as if the other possesses traits included in the stereotype. Stereotypes are often expressed implicitly in the behavior of persons who explicitly disavow the stereotype. The authors remind us that distraction increases implicit effects, attention to source of implicit effect reduces the effect, and that recall of implicit cue decreases implicit effect.

Wild, H., Barrett, S., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000): young children may master the use of sex-stereotyped social and cultural cues to gender earlier than they master the use of facial structure cues to sex. *Notes on Early Media Exposure.*

Blthoff I, (2009): Participants asked to ignore identity and base their responses solely on the sex appearance of the faces, showing that participants were slower and less correct for sex-changed than for unchanged familiar faces while those differences did not appear for unfamiliar faces. These results indicate that sex and identity are not independent as participants could not ignore identity information while doing a sex categorization task. This is in the context of visual perception; the current work would supplement it with speech perception.

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013): Participants categorized faces more accurately and more quickly by sex (670 ms) than race (712 ms), and that prior work suggests that FFA is active just 200 ms after seeing a face, suggesting that differences in race and sex are parsed early in visual perception.

8 References

1. Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, *113*, 544-552.

2. Andersen, E. S. (1990). *Speaking with style: The sociolinguistic skills of children.* London: Routledge.

3. Andersen, E. S., & Johnson, C. E. (1973). Modifications in the speech of an eight-year-old as a reflection of age of listener. *Stanford Occasional Papers in Linguistics*, 3, 149-160.

4. Andersen, E. (1979). Register variation in young children's role-play speech. In *Proceedings of the Society for Research in Child Development.*

5. Aubrey, J. S., & Harrison, K. (2004). The gender-role content of children's favorite television programs and its links to their gender-related perceptions. *Media Psychology*, 6(2), 111-146.

6. Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology, 71(2)*, 230.

7. Bellinger, D. C., & Gleason, J. B. (1982). Sex differences in parental directives to young children. *Sex Roles*, 8(11), 1123-1139.

8. Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2008). Online expectations for verbal arguments conditional on event knowledge. *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2220-2225). Austin, TX: Cognitive Science Society.

9. Brouwer, D., & de Haan, D. (1987). *Women's language, socialization and self-image (Vol. 4).* Foris Publications.

10. Blthoff I, (2009). Sex categorization is influenced by facial information about identity. *Perception*, *38*, ECVP Abstract Supplement, 78.

11. Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

12. Comstock, G., & Scharrer, E. The Use of Television and Other Screen Media. in Singer, D. G., & Singer, J. L. (Eds.). (2011). *Handbook of Children and the Media.* Sage publications.

13. Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel patterns in fusiform face area differentiate faces by sex and race. *PloS one, 8(7)*, e69684.

14. Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *108*, 633-664.

15. Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*, 264-285.

16. Creel, S. C. (2012). Preschoolers' Use of Talker Information in On-Line Comprehension. *Child development*, *83*, 2042-2056.

17. Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 498.

18. Fellowes, J. M., Remez, R. E., & Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics*, *59*, 839-849.

19. Gleason, J. B. (1973). Code Switching in Children's Language. In Moore, T. (Ed.). *Cognitive Development and the Acquisition of Language (Academic Press).* pp. 169-167.

20. Gleason, J. B. (1975). Fathers and Other Strangers: Men's speech to Young Children. *26th Annual Georgetown University Roundtable (Georgetown University Press).* pp. 289-97.

21. Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166-1183.

22. Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, *105*, 251.

23. Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review, 102(1),* 4.

24. Greif, E. B. (1980). Sex differences in parent-child conversations. *Women's Studies International Quarterly*, 3(2), 253-258.

25. Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*, 105-115.

26. Johnson, K. A., Strand, E. A., & D'imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, *24*, 359-384.

27. Johnson, K. A. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 363-389). Oxford: Blackwell.

28. Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of phonetics*, *34*, 485-499.

29. Kempson, R. (2001). Pragmatics: Language and communication. In M. Aronoff & J. Rees-Miller (Eds.), *Handbook of linguistics.* Malden, MA: Blackwell.

30. Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29*, 98-104.

31. Liberman, A. (1970). The grammars of speech and language. *Cognitive Psychology*, *1*, 301-323.

32. Maccoby, E., & Jacklin, C. (1974). *The Psychology of Sex Differences.* Stanford: Stanford University Press.

33. McGillicuddy-De Lisi, A. V. & De Lisi, R. (Eds.). (2002). *Biology, Society, and Behavior: The Development of Sex Differences in Behavior.* Westport: Greenwood Publishing Group. Has even more gender stereotypes work and a variety of other perspectives on gender behavior.

34. Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, *6*, 313-330.

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

36. Nygaard, L. C. (2005). Perceptual integration of linguistic and non-linguistic properties of speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 390-413). Oxford: Blackwell.

37. Rideout, V. J., Vandewater, E. A., & Wartella, E. A. (2003). *Zero to six: electronic media in the lives of infants, toddlers and preschoolers.* Henry J. Kaiser Family Foundation.

38. Rideout, V., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers and their parents.* Henry J. Kaiser Family Foundation.

39. Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309.

40. Patterson, M. L. & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, *81*, 93-115.

41. Perry, J. (1997). Indexicals and demonstratives. In C. Wright & R. Hale (Eds.), *Companion to the philosophy of language* (pp. 586-612). Oxford: Blackwell.

42. Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*, 2988-2998.

43. Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*, 175-184.

44. Philips, S. U., Steele, S., & Tanz, C. (Eds.). (1987). *Language, gender, and sex in comparative perspective* (No. 4). Cambridge University Press.

45. Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 651.

46. Rideout, V. J., Vandewater, E. A., & Wartella, E. A. (2003). *Zero to six: electronic media in the lives of infants, toddlers and preschoolers.* Henry J. Kaiser Family Foundation.

47. Rideout, V., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers and their parents.* Henry J. Kaiser Family Foundation.

48. Singh, S., & Murry, T. (1978). Multidimensional classification of normal voice qualities. *Journal of the Acoustical Society of America*, *64*, 81-87. Retrieved July 23, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/712004

49. Sachs, J. (1987). Preschool boys? and girls? language use in pretend play. *Language, gender, and sex in comparative perspective*, 178-188.

50. Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive psychology*, *49*, 238-299.

51. Van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*, 580-591.

52. Van Lancker, D., Kreiman, J., & Wickens, T. D. (1985). Familiar voice recognition: Patterns and parameters, part II: Recognition of rate-altered voices. Ĵournal of Phonetics, *13*, 39-52.

53. Wild, H., Barrett, S., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000). Recognition and Sex Categorization of Adults' and Children's Faces: Examining Performance in the Absence of Sex-Stereotyped Cues. *Journal of Experimental Child Psychology*, *77*, 269-291.

54. Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37-63.