ABSTRACT

Title of dissertation: An Affiliative Model of Early Lexical Learning

Alayo Rachael Tripp, Doctor of Philosophy, 2019

Dissertation directed by: Professor Naomi Feldman
Department of Linguistics

Professor William Idsardi
Department of Linguistics

In defining the language acquisition problem, traditional models abstract away effects of variability, defining the learner as acquiring a single language variety, which is spoken homogeneously by their speech community. However, infants are exposed to as many unique varieties of speech as they are speakers. Adult sociolinguistic competence is also characterized by the capacity to employ and interpret non-phonological linguistic distinctions which are associated with different social groups, including 'code-switching' or 'style-shifting' between languages and speech registers.

This dissertation presents a model of infant lexical acquisition which assumes that learners monitor linguistic sources for variation in reliability. This model is adapted from Shafto, Eaves, Navarro, and Perfors (2012) which the authors used to describe the behavior of preschool children in selecting sources to learn labels from in K. Corriveau and Harris (2009) and M. Corriveau and Harris (2009). I show that this probabilistic model effectively simulates two experiments from the literature on preverbal infants' perception of labeling, Rost and McMurray (2009) and Koenig and Echols (2003). Evidence suggests that the receptiveness of preverbal infants to novel lexical items is correlated with

infant beliefs regarding the informant's knowledgeability and social group membership. These simulations demonstrate that language learners may well be recruiting processes of epistemic trust to guide lexical acquisition much earlier than previously suggested.

We should therefore expect even very young listeners to respond differently to dialects not solely as a function of exposure, but also as a function of attitudes towards the speech determined by the quality of that exposure. Developmental differences between populations in attention to non-linguistic affiliative cues are therefore expected to emerge early and have significant effects on language outcomes. Measures of online language proficiency may be vulnerable to significant bias owing to the activation of sociolinguistic biases in the presentation of test items. Differences in the breadth or specificity of listener preferences for speakers in turn predict differences in task complexity for learners of standard and non-standard dialects. A new research program in early sociophonetic perception, uniting accounts of selective trust with language learning has the potential to deepen understanding of both typical and disordered language development.

AN AFFILIATIVE MODEL OF EARLY LEXICAL LEARNING

by

Alayo Rachael Tripp

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
2019

Advisory Committee:
Chair: Associate Professor Naomi H. Feldman
Professor William Idsardi
Professor Jeffrey Lidz
Assistant Professor Yi Ting Huang
Professor Jan Edwards

Acknowledgments

As an undergraduate, the Linguistics department at the University of Maryland is where I first discovered the field of linguistics. As a graduate student, it is where I discovered myself as a language scientist. It would be impossible for me to properly express gratitude in this space to all the people who helped me on this journey.

I have had the guidance of two phenomenal mentors, Naomi Feldman and William Idsardi. They and the rest of my committee have struck a wonderful balance between nurturing my creative impulses and demanding I learn how to pace myself. I am grateful for their criticism and their encouragement, and the many lessons they have implicitly given me in mentorship. I hope that I can one day live up to their example.

I owe special thanks to Yi Ting Huang and Jonathan Beier for igniting my pursuit of this dissertation topic, and to Jan Edwards and Jeff Lidz for their helpful feedback. I also want to recognize the department administrator Kim Kwok for her support navigating all manner of administrative exigencies over the years.

I am grateful for the many students across the Language Science community, especially the computational modeling lab group, who helped me to develop my ideas and refine their presentation. Special thanks to Alexander Sushenov, who worked with me as an undergraduate to demystify the source evaluation model.

Finally, I am extraordinarily grateful to all my friends and family who never showed anything but the greatest faith that I would succeed.

Contents

List of Figures

List of Tables

Chapter 1

Learning about Words in the Context of Affiliative Cues

1.1 Learning Words from Sources

The goal of this work is to investigate what effects the processes that infants use to categorize informants may have on the phonetic perception and lexical acquisition of pre-verbal infants. How do infants form beliefs about the relative quality of speech sources, and what is the role of these beliefs in shaping infant attention to the input? I will begin in this chapter by defining the word learning problem, and outline how existing models abstract away effects of source variability.

If infants are not using judgments of informant quality to guide their acquisition of words, we should not expect a model positing that perceptual categorization of informants directly influences word learning to accurately capture their behavior. However, demonstrating that such a model is adequate to explain infant behavior on word learning tasks obviously also does not discredit alternative theories on the subject. However, I submit that a model of language acquisition which recruits domain general epistemic trust provides a comprehensive and parsimonious account of early word learning in the presence of variation in language forms, with potentially massive implications for the study of language acquisition.

The claims made in this dissertation, that infant knowledge about informant quality may affect word learning, should be viewed as compelling for three reasons. First, there is a large body of work which suggests that judgments of an informant's value as both a linguistic and a non-linguistic source of information are tightly linked (Schachner and Hannon, 2011; Kinzler and Dautel, 2012; Z. Liberman, Woodward, and Kinzler, 2017). This is expressed with reference to concepts such as perception of an "in-group" which has characteristic linguistic and non-linguistic features and behaviors. Studies of the ontogeny of attitudes towards in-group and out-group members (i.e. David Buttelmann and Böhm, 2014; Mahajan and Wynn, 2012) are predicated on the accepted assumption that socially motivated cognitive grouping is part of human behavior and likely emerges quite early.

Second, there is an equivalent amount of literature demonstrating that preverbal infants, including newborns, use non-linguistic observations to evaluate social partners (Akhtar and Gernsbacher, 2008; Coulon, Guellai, and Streri, 2011; Maurer and Werker, 2014; Cirelli, Wan, and Trainor, 2016) and that the processing of indexical and referential information are neurally functionally integrated in adults (for a review see: Patricia K Kuhl, 2011).

Third, the observation that linguistics has historically given short shrift to matters of social practice and the role of social power in language use has been made many times before (Eckert, 2008; Group" et al., 2009; Legare and Harris, 2016). This dissertation defines a unified framework in which to view research on early language across within and between diverse populations. Further, it calls for additional work to refine theories about infant responses to speech stimuli by making explicit assumptions about the social value of the speech being used in a given investigation. In particular, I am concerned with studies which explore how or whether infants interpret the referential content of speech in the face of language variation. I argue that existing theories of sensitivity to language variation and consequently, of early lexical acquisition, often fail to explicitly ground their characterizations of infant linguistic knowledge in a context of infant social cognition and adequately specify the scope of their investigation.

For example, Schmale, Cristia, Seidl, and Johnson (2010) familiarized 9 and 12 month old children from two English speaking regions with words spoken either in their own or the opposite regional language variety. They were then tested with passages that did or did not contain the familiarized words. Only the 12 month olds showed a significant difference in looking times between hearing novel and familiar word spoken in the unfamiliar dialect. The conclusion that by 12 months, children can segment words from continuous speech across minimally different dialectal accents takes for granted that one of the presented dialects was familiar, neglecting the social implications of an informant using a familiar dialect. Butler, Floccia, Goslin, and Panneton (2011) makes this argument, presenting evidence that 7 month olds differentiated Welsh English from their home dialect of West Country regional English, but did not discriminate between two unfamiliar accents (Welsh and Scottish), despite these dialects being rated as being equivalently similar by American listeners.

The usage of American listeners reflects the belief that their sociocultural naivety is loosely comparable to the social knowledge of the West Country learning infants. However, making this assumption about the role of the infant's social knowledge explicit supposes that at this age infant speech processing is not influenced by non-linguistic judgements about informant quality or affiliation. Supposing allocations of epistemic trust are made independently of linguistic judgments, we should not expect learning about an informant's competence in one domain to affect the infant's perception of that informant's performance in the other. The argument that infant word learning is impacted by social perception should therefore be considered an uncontroversial claim which does not pose a challenge to existing theories. This dissertation sets forth a framework within which to ground theories of early lexical acquisition by making explicit assumptions about social power, outlining a research program to more precisely delineate the role of social perception in early word learning.

This chapter contains a review of the literature on infant discrimination of informants motivating my model of infant word learning from sources which vary in reliability. This model, adapted from Shafto et al. (2012) is presented in the subsequent chapter.

In later chapters I will apply this probabilistic model to simulate two experiments from the literature on infant perception of labeling, Rost and McMurray (2009) and Koenig and Echols (2003). While evidence shows that selective trust of informants guides the preferences of pre-school age children (K. Corriveau and Harris, 2009;M. Corriveau and Harris, 2009), my simulations will demonstrate that language learners may well be recruiting similar processes of informant evaluation to guide word learning in infancy.

1.1.1 The Dialect Learning Problem

Language acquisition can be thought of as a statistical inference problem. Given linguistic input, a learner must recover the structure of their language. However, this model lacks the explanatory power to describe how language learners come to recognize labels produced by speakers of varieties which do *not* represent their target dialect.

I use “target dialect” here to denote any necessary collection of sociolects, ethnolects and other socially-marked forms of speech. The phrase “non-target dialect” then refers to all other socially-marked variants of that language.

Importantly, classifications of accent, dialect and language specify the parameters for dividing both speech *and speakers* into categories which are *functionally* homogenous.

The child's emergent sensitivity or insensitivity to typological contrasts tells us then, not only about how they classify speech, but about how they prefer to classify speakers. Acquisition of the target dialect may therefore be conceptualized as proceeding from the solution of a simpler problem: successfully identifying its speakers.

In models of language acquisition the traditional assumption is of a uniformly reliable input signal describing the behavior of a single linguistic community. Instead, I intend to explore the consequences for models of word learning of characterizing the input as distributed over sources which vary in reliability and which are representative of multiple linguistic communities. I will show evidence that infant word learning is socially biased - infants use non-linguistic cues to social group membership, or *affiliative cues* to predict both affiliative linguistic and non-linguistic behavior.

If latent group membership predicts knowledgeability, and knowledgeability predicts the referential content of an informant's speech, any perception which implies that the speaker holds membership in a group will also impact the perception of the referential content of their speech. In other words, perception of affiliative cues is expected to have an a priori impact on beliefs about the reliability of both linguistic and non-linguistic informant behavior. Affiliative cues will therefore necessarily affect the linguistic behavior of the listener by exerting an influence on their beliefs about an observed informant's production and perception patterns before that informant speaks.

I will begin by describing the literature on language users' beliefs about affiliation and reliability in both linguistic and non-linguistic contexts. Next I will outline how observing an informant's linguistic behavior shapes infant expectations about both affiliation and reliability. I will then explore the implications of these findings for accounts of early word learning. To the extent that group membership predicts knowledgeability, and knowledgeability predicts the referential content of an informant's speech, any perception which implies that the speaker holds membership in a group may also impact the perception of the referential content of their speech.

1.2 Speech and Membership in Social Groups

Adults interpret language in a social context, showing sensitivity to non-linguistic social cues in both linguistic perception and production. In acquisition, therefore, we may conclude that the adoption of productive linguistic behavior associated with specific native-language using social groups indicates that the language learner is sensitive to variation between the speech patterns of speakers who are and are not members of those groups. In other words, children eventually adopt production patterns matching their own social groups, demonstrating that the language learner is sensitive to linguistic division *among native speakers*.

Young children have the ability to identify speakers as individuals, and by socially defined groups, including gender. However, it is not clear how early these abilities emerge, and what role they play in lexical acquisition.

If discerning native from non-native language patterns is the process of deciding which language patterns merit mimicry, the adoption of more specialized dialectical speech suggests a decision process discriminating which of the valid native language patterns they are exposed to merit even closer mimicry than other speakers. This line of reasoning demands we ground the problem of linguistic source evaluation in what is already known about how infants evaluate sources for non-linguistic purposes. In other words, how does infant perception of speech develop in relation to non-linguistic judgments about speakers?

To successfully acquire their native language, culture and customs, learners must be able to distinguish between behavior patterns (linguistic and non-linguistic) which are representative of their social group, and those which are not. Nevertheless, a developing preference for imitating the acoustic patterns of, for example, either male or female speakers does not prevent children from acquiring linguistic units from the dispreferred category of speakers. Therefore, acquiring a native language necessarily demands that the learner distinguish between speakers who represent different versions of the native language - that is, speakers who are representative of speech sub-communities.

In the next section I will argue that infant behavior suggests both the perception of linguistic sub-communities and listener beliefs about their own membership in these groups. I argue that acquisition of sociolinguistic variation necessarily hinges on beliefs about social groups (i.e. masculine and feminine speakers) and their characteristic traits (i.e. masculine or feminine physical features) and behaviors (i.e. masculine or feminine speech). Development of atypical language skills is therefore predicted to consistently follow atypical beliefs about sources.

1.2.1 The Acquisition of Sociolinguistic Variation in Production

The social construct of gender provides a useful example of affiliative language learning. Gendered sociolinguistic norms vary both within and across languages, and adult interactions are colored by both social and linguistic biases.

One type of account for how these biases are acquired draws on an assumption that attention to physical features which are correlated with social judgements (i.e. vocal tract length, skin color or eye shape) are sufficient to explain the development of associated beliefs about linguistic behaviors. However, this kind of account is highly specified, requiring separate accounts for each non-linguistic percept shown to impact linguistic perception in any modality. This type of account is also generally ambiguous regarding the question of whether the feature in question is universally attended to in speech perception, or if the behavior of discriminating it is specific to the population being studied.

Well before physical sexual dimorphism occurs, children begin showing gendered features in their speech production (P. Foulkes, Docherty, and Watt, 1999; Perry, Ohde, and

Ashmead, 2001; Simpson, 2009). Even supposing that acoustic cues provide sufficient basis for categorizing speech patterns by gender, acquisition of gendered speech also reflects the acquisition of affiliative beliefs about the child's own gender.

If the child's intake were constructed by sampling without any attention to source, we would expect a child with consistent exposure to two gendered linguistic patterns to simply acquire an idiolect which is a mixture of the two. On the other hand if the learner were excessively sensitive to source variation, they might acquire separate dialectal representations for each interlocutor they encounter.

Contrastively, the affiliative model defines a framework for interpreting speech in light of beliefs about the group membership of both speaker and listener, defining the relevance of non-linguistic features in a linguistic context as dependent on properties of the specific language users, rather than universal properties of the language faculty. This framework easily accommodates variation in learners across different cultures and modalities.

A socially grounded language acquisition process demands children discern the relative value of speakers as linguistic models for latent groups. In effect, children must be selectively attending to the linguistic behavior of those speakers who reflect the listener's own emerging linguistic identity. This pattern is readily apparent in the adult language user's ultimate productive and perceptual competence regarding socially marked speech forms.

I will argue that children learn to recognize the latent groups which informants belong to, learning distinct standards of linguistic behavior for each. Pre-existing beliefs about the characteristics of groups, and about the membership of speakers, are therefore predicted to produce distinct learning outcomes, even for children with identical exposure to speech.

Apart from gender, language users employ linguistic features which mark other aspects of their social identity, including ethnicity and social class (Paul Foulkes and Docherty, 2006). If a learner treated all speakers as equally reliable, then we should expect children to only acquire such distinctly socially marked speech patterns according to rigid segregation between caretakers. However, frequency of exposure does not appear to account for the acquisition of a gender-marked speech patterns emerging early in child speech (Ladegaard 2003). Further, children whose parents speak their target language with a foreign accent naturally acquire a native accent, and children who are raised by a caregiver of a different gender nevertheless acquire the speech patterns of their own gender. Floccia, Delle Luche, Durrant, Butler, and Goslin (2012) showed that 20-month old infants more often recognized words in the rhotic accent of their community, even when one or both parents spoke with non-rhotic accents. Social categorization, rather than frequency of exposure, appears to be the greatest influencing factor in children's linguistic representations at an early age.

In this section I broadly review evidence that infants make judgments about people and decide which informants to preferentially learn details from in both linguistic and non-linguistic contexts. What role does the formation of social knowledge play in shaping the development of the lexicon? Even assuming that informant reliability and linguistic affiliation are independently determined, evidence suggests that infant inferences about reliability and social group membership greatly influence expectations for linguistic behavior.

1.2.2 Language, Affiliation and Infant Social Partner Selection

Very young children show evidence that suggests emerging social preferences, demonstrating cognitive biases for some groups over others. Neonates prefer to view images of their mother's face and listen to her voice (DeCasper and Fifer, 2013, 4448), prefer to view faces which match their primary caregiver in gender (Quinn, Yahr, Kuhn, Slater, and Pascalis, 2002) and the faces of racial in-group members (D. J. Kelly et al., 2005, Bar-Haim, Ziv, Lamy, and Hodes, 2006). There is further evidence that infants use may their sensitivity to physical features of informants for categorization tasks. As early as 10 months of age, infants show sensitivity to both gendered and racialized features, and a willingness to use statistical regularities in the presentation of these features to categorize objects which are associated with them in pre-test (Gary D. Levy and Haaf, 1994; G. D. Levy, 2003). They also prefer to listen to their native language (Mehler et al., 1988; Pascalis, de Schonen, Morton, Deruelle, and Fabre-Grenet, 1995; Nazzi, Bertoncini, and Mehler, 1998; Moon, Cooper, and Fifer, 1993).

In chapter 5 I will review the literature on infant beliefs about social groups in more depth. In the next section I review the literature describing infant beliefs about the quality of informants as models for both linguistic and non-linguistic tasks.

.

1.3 Quantifying Informant Quality

Suppose a listener has some pre-existing knowledge of their language structure. We might describe our listener as judging an informant's linguistic quality based on known statistical relationships between the acoustics and the phonological categories. A successful language learner must acquire the ability to contrast data from speakers who are more or less accurate. This process of identifying which informants are verifiably valuable sources of information is the problem of epistemic trust. Alternatively, there is evidence that infants tend to simply assume informant behavior will be highly informative. This theory is called natural pedagogy.

In this section, I detail the evidence in the literature of how learners make judgments about the reliability of informants. In the next chapter I will introduce a model which unites these two accounts of source selection, natural pedagogy and epistemic trust, predicting that confidence about speaker reliability controls word learning behavior (Shafto et al., 2012).

1.3.1 Natural Pedagogy

Children and infants show great sensitivity to ostensive cues, which convey generalizable knowledge rather than factual information (Csibra and Gergely, 2006). Infants also prefer to attend to details supplied by informants in ostensive contexts, where the intent to communicate is evident. Research investigating children's non-linguistic learning through social interaction suggests that they more readily learn when information is generated in a pedagogical setting (Gergely, Bekkering, and Kiraly, 2002; Gergely, Egyed, and Kiraly, 2007). Csibra and Gergely (2009) proposed that children's sensitivity to ostensive cues is the product of an inclination to identify and understand acts as pedagogical.

For example, children often over-imitate adult models in non-linguistic tasks, reproducing even causally irrelevant gestures (Lyons, Young, and Keil, 2007; McGuigan and Whiten, 2009; Kenward, Karlsson, and Persson, 2011; Nielsen and Blank, 2011). Effectively, children often appear to assume that informants will both be knowledgeable and provide representative data, especially in the presence of social information which supports this inference.

There is evidence that in the first year of life, infants are already recruiting social signals that bias the ways they attend to objects. For example, 9-month-old infants preferentially encode the identity of novel objects when they have seen a model point at them, as opposed to when the informant reaches for the object (Yoon, Johnson, and Csibra, 2008), and imitate simple actions using novel toys both immediately and after a delay

(Meltzoff, 1988). Moll, Richter, Carpenter, and Tomasello (2008) showed that when an experimenter excitedly made an ambiguous request for an object, 14-month old infants preferred to select the object which had been the subject of previous shared joint attentional experiences with that experimenter.

Applying this account of informant selection to language acquisition would predict that language learners begin with the assumption that speakers are reliable before making an inference about linguistic categories. Listeners would then be expected to re-evaluate this belief about the speaker only when confronted with evidence that contradicts it. This account predicts that infants use epistemic trust processes to learn selective avoidance of speakers who are less helpful.

1.3.2 Epistemic Trust

Another account of how children learn about informants in non- linguistic tasks shows that infants infer the knowledgeability of informants from evidence, both learning to selectively attend to informants who are more useful and learning to avoid those who are less so.

Independent of variation in non-linguistic ostensive cues, young children's selective trust in informants has been shown to be manipulated with the child's attachment to the informant (Harris and Corriveau, 2011), their observed reliability (K. Corriveau and Harris, 2009; Koenig and Harris, 2005) but also notably with the native accentedness of the speaker (Kinzler, Corriveau, and Harris, 2005).

However, this ability to track the reliability of an informant emerges much earlier in life. At 8 months old, young infants can track the reliability of a novel informant, and distinguish between similar cues (i.e the directional gaze of a pictured face) from informants who have previously proven either reliable or unreliable for indicating the location where an object will appear (Tummeltshammer, Wu, Sobel, and Kirkham, 2014).

Xiao et al. (2018) demonstrated that 7-month-old infants also prefer to attend to speakers whose gaze is consistently reliable. However, when presented with two speakers who both exhibit only 50% reliability, the infants prefer to attend to the speaker who shares their own race. This suggests that infants are integrating new information about epistemic reliability with their prior knowledge of speakers, and attending to those speaker attributes which they a priori associate with knowledgeability.

While children and infants often over-imitate, their propensity to treat novel informants as reliable may be influenced by epistemic observations that an informant is not reliable. For example, At 14 months old, infants are significantly more likely to to imitate a model who they have seen using objects correctly (i.e. putting sunglasses on their face) as opposed to a model who they have witnessed using objects incorrectly (i.e. putting sunglasses on their foot) (Zmyj, Buttelmann, Carpenter, and Daum, 2010).

Likewise, in Poulin Dubois, Brooker, and Polonia (2011) 14-month olds who watched an experimenter turn on a touch light using her forehead were more likely to imitate this causally irrelevant behavior if they had previously witnessed that experimenter providing reliable affective cues. Those infants who had previously viewed the experimenter showing unreliable affective cues were less likely to over-imitate, instead using their hands to turn on the touch light. Infant representations of the informant therefore reflected previous observations regarding their value as models for imitation (Poulin Dubois et al., 2011).

1.3.3 Selection of Linguistic Informants

Both of these accounts predict that infants reassess their beliefs about informants' knowledgeability, valuing informants in accordance with the availability of evidence which supports their predictions. While there is a lot of evidence that preverbal infants epistemically infer the quality of informants in non-linguistic tasks, I am interested in how these processes of selective trust may impact early word learning.

If the knowledgeability of linguistic and non-linguistic informants are assessed in wholly distinct ways, we should expect to find a double dissociation between learning about competence in each domain, however, infant judgments of an informant's value as both a linguistic source of information greatly influence their beliefs about the informant's value as a non-linguistic source of information and vice versa.

There is substantial evidence that before the second year of life, infants have already formed beliefs about language and affiliation. Z. Liberman et al. (2017) demonstrated that 9-month olds who view two people using different languages are surprised if they subsequently display affiliative behavior. Similarly, if they view these people using the same language, infants are surprised to witness subsequent disaffiliation (Z. Liberman et al., 2017). These results suggest that not only do infants preferentially interact with speakers of their own native language, they also expect others to behave likewise.

At 14 months, infants also imitate a model more faithfully in a non-linguistic task, given that the model speaks their native language as opposed to a foreign one (D. Buttelmann, Zmyj, Daum, and Carpenter, 2013, 2). From an early age, the social signal of a shared language appears to serve in and of itself as an ostensive cue to attend to non-linguistic information. But in what ways do infants discriminate between language users specifically as sources of linguistic information? I wish to investigate how beliefs about reliability and ostensive cues impact assessments of informants as sources of linguistic information in preverbal children.

Brooker and Poulin Dubois (2013) investigated infants' attention to adults' linguistic accuracy, demonstrating that at 18 months, infants not only track an informant's reliability in labeling familiar objects, but use these observations to make judgments about the value of the informant as a model for both linguistic and non-linguistic behavior. Infants exposed to an inaccurate labeler were both less likely to perform well on a subsequent word-learning task and less likely to imitate the behavior of that labeler in a non-linguistic task. Given infants' proven ability to track the competence of informants, it is possible that these processes exert an influence earlier in the lexical acquisition process.

As infants age, their preferences for native language informants grow stronger - for example, 6-month-old infants prefer to accept a toy from a speaker of their own language (Kinzler, Dupoux, and Spelke, 2007). This native language preference persists as children continue to learn about social groups. Kinzler, Shutts, DeJesus, and Spelke (2009) presented five year old children with the images of two children, each paired with a voice recording either in English or French. Children were asked to indicate which child they would prefer to be friends with. In a second experiment, the provided linguistic contrast was between native and foreign accented English. Participants preferred to select children of the same race over different race children when the prospective friends were silent, but when exposed to the speech contrasts expressed a stronger preference for children who speak their native language variety with a native accent when allowed to observe the children's speech.

The basic idea is that speakers who produce data that the listener can identify as native are considered preferable sources of information to both speakers who are linguistically similar, but non-native, (i.e. accented) and speakers whose speech is from another language, and these preferences emerge in infancy. As children acquire their first language, judgments about linguistic group membership exert an increasing influence on the child's selection of social partners. For example, 6-month-old infants look longer at a silent model who has previously spoken in the child's native language. Consistent with this, when silently presented with a toy offered by a person who previously spoke either English or French, 10-month old French speaking infants prefer to accept the toy from the person who spoke French (Kinzler et al., 2007).

Children's perceptual linguistic knowledge appears to reflect distinctions between foreign and native dialects beginning even earlier in development. Polka and Sundara (2012) show that Canadian French-learning 8-month old infants could segment words in European French but not in English. We easily intuit that a Canadian-French learning child should find utterances from a European French speaker of higher utility than those from an English speaker. However, setting aside the mechanics by which the segmentation may take place, I wish to emphasize the implicit hierarchical social consequences.

The perception that two individuals share a language appears to motivate young infants to expect these individuals to share a social group. Observing similar linguistic and similar non-linguistic behavior both appear to function as cues to the infant that the informants belong to the same linguistically defined group. Rather than generalizing variations in speech and communication patterns to all speakers, or even to all speakers of a language, listeners' generalizations describe contrasting subsets of speakers.

Child processing of different varieties of in-group speech must therefore rely on some set of mechanisms for evaluating the relative quality of native informants and categorizing them. Within the context of a target dialect belonging to a target language, the value of a particular informant may therefore be modeled with a measure of relative utility. In the next section I discuss an experiment from Kinzler and Dautel (2012) which investigates children's beliefs about language and race as markers of social group membership. The findings demonstrate that depending on their own social identity, children have different beliefs about which features are affiliative. If beliefs about non-linguistically defined groups impact the selection of linguistic models then we should expect distinct developmental trajectories for children of different backgrounds even controlling for exposure to the target dialect.

1.4 The Impact of Differences in Social Group Membership on Perception

Supposing that beliefs about affiliation affect an infant's language acquisition, infants with different beliefs about social groups should be expected to perform differently on tests of language. Children who are aware of their membership in distinct social groups could therefore be accurately described as having distinct language acquisition processes, even supposing that their target dialect is identical.

Kinzler and Dautel (2012) conducted a series of experiments investigating children's beliefs about language and race as markers of social group membership. Children were shown an image of a child, each paired with a recording of speech in either English or French. Children were then asked to match each child to one of two adults, indicating who each target child would grow up to be. One of the adults was the same race as the target child but spoke another language, and the other spoke the same language as the target child but was of a different race. (Kinzler and Dautel, 2012).

Kinzler and Dautel (2012) found that at 5-6 years old, European American children from different backgrounds selected the language match, despite the implication that this meant the child would change races. Nine to 10 year old European American children selected the race match, suggesting that they understood race to be an stable cue to identity and group membership. African-American children, however, at 5-6 years old already selected the race match, suggesting they have the expectation that race is a more stable category than language.

This study illustrates that recruitment of cues to group identity is different across social populations. To the extent that the Black children showed a more adult-like grasp of social categories, what consequences should we expect for their word learning? If we assume that this distinction in social perception is unremarkable from a linguistic standpoint, we should expect that learners from these populations who are supposedly learning the same language will perform comparably on assessments of language development.

However, children who come from lower socioeconomic strata (SES) and language minority homes have distinct language trajectories (Brooks-Gunn, Rouse, and McLanahan, 2007;Huttenlocher, Waterfall, Vasilyeva, Vevea, and Hedges, 2010). Washington, Craig, and Kushmaul (1998) found that within a population of African-American children, both SES and gender predicted the usage of marked dialectal forms. In order to accurately describe the variation in developmental trajectories of children who speak the same language, in addition to studying environmental differences, it is necessary to assess how perception of these social categories emerges, and how preferences for certain kinds of informants shape the formation of lexical representations.

1.5 Outline of Dissertation

This dissertation explores the role of non-linguistic social cognition on attention to lexical input, and the limitations of word learning models which do not control for infants' speaker preferences. Beginning with the assumption that there are social groups which vary in relative reliability, evidence which suggests that an informant belongs to a group which is proportionately more reliable therefore lends proportionally more weight to the hypothesis that the labels they provide are accurate. However, direct evidence of informant accuracy about a label obviates attention to social group membership.

Models of language acquisition which do not account for variation in paralinguistic cues essentially idealize that all sources are both equally reliable, and indistinguishable on dimensions apart from those which encode referential content. Such models therefore predict that neither linguistic nor non-linguistic behavior will be affected by contrasts in speaker identity, or by the presence of other passive listeners. A more effective model of lexical acquisition must account for how beliefs about informants and their speech appear to be a priori affected by “non-linguistic” social cues.

In this work, I depart from the traditional assumption of a uniformly reliable input signal, instead characterizing the input as distributed over multiple sources varying in reliability. Shafto and Goodman (2008) introduced a model of reasoning about knowledgeability and helpfulness by informants and learners, which has been shown to capture the behavior of test subjects across a variety of conditions (Gweon, Pelton, and Schulz, 2011; Shafto, Goodman, and Griffiths, 2014). It is our goal to apply this model to describe the behavior of infants in labeling tasks, exploring the role of judgments about speaker quality in early word learning.

To this end, I adapt a model presented in Shafto et al., 2012. Some key differences are that the proposed model requires the learner to make inferences about potentially ambiguous data, and rather than contrasting informants who are helpful with those who actively hinder the learner, I will assume all informants are minimally helpful, with some being significantly more helpful than others. This approach guarantees success for any learner who can identify a knowledgeable informant.

I will use this model to reproduce the pattern of results from a number of experiments using a violation-of-expectation methodology that attempts to induce word learning in preverbal infants (Stager and Werker, 1997). I will then extend our model to describe how interpreting behavior as affiliative may provide an alternate explanation for these findings, and discuss the implications for experimental work on early word learning.

Lastly, with an eye to the social information infants recruit when making partner selections, we will critically examine vocabulary as a measure of linguistic aptitude and development for language learners, especially those whose target dialects represent minority communities.

Chapter 2

Modeling Early Lexical Learning in a Social Context

Word learning is an example of a domain where a child must learn to resolve linguistic information from multiple contrasting informants, and provides a useful illustration of how epistemic trust interacts with existing linguistic knowledge to shape the acquisition process. Supposing that the utility of the linguistic data supplied is not uniformly consistent across informants, we can then describe this utility as a probabilistically defined function of the informants' qualities. The model introduced in this section will allow us to explore the constraints on a learner who is sampling from sources who exhibit trustworthy and/or pedagogical behavior to infer the correct linguistic categories under uncertainty about speaker qualities.

This model is an adaptation of the model from Eaves and Shafto (2017), which simulated the behavior of preschool age children on a labeling task. Following M. Corriveau and Harris (2009), the authors demonstrated that 3 and 4 year old children's preference to answer the question "Which is the [novel object label]?" with information given by an informant who had previously agreed with a majority, over information from an informant who had dissented, implies that children are attending to the credibility of linguistic sources when learning words. I hypothesize that children may be relying on similar socially motivated inferences to guide their processing of speech at a much earlier age.

The adapted model presented here is distinct in that it requires the learner to make inferences about potentially ambiguous data, and assumes all informants are minimally helpful, rather than contrasting informants who are helpful with those who actively hinder the learner. To model the dialect learning problem, I assume that all informants speak the target language, and the learner's task is determining which informants are more helpful for the purpose of learning a target dialect.

In this chapter we will begin by outlining a dialect learning inference problem. Focusing on a perspective from phonetics and phonology, we will demonstrate the comparative utility for word learning of linguistic input which is specifically pedagogical to that which is merely epistemically trustworthy.

Supposing that the utility of the linguistic information supplied is not uniformly consistent across informants, we can then describe this measure as a function of beliefs about informants' qualities. In the next section we will outline learning object labels as an inference problem. Focusing on a perspective from phonetics and phonology, with a series of simulations we will demonstrate the comparative utility for word learning of linguistic input which is specifically pedagogical to that which is epistemically trustworthy.

2.1 The Dialect Learning Problem

Preferences for informants who exhibit one type of linguistic behavior over another functionally reflect infant beliefs about both the relative value of linguistic representations themselves and the informants who appear to employ them. By virtue of attentional preferences favoring some linguistic informants over others, infants demonstrate distinct expectations of informants depending on observations of information access (Koenig and Echols, 2003), speaker identity and epistemic reliability (Poulin-Dubois and Chow, 2009), and group membership (Kinzler et al., 2007).

To the extent that judgments about categories of informant may affect a learner's beliefs about the epistemic value of an informant's linguistic data, we should expect both linguistic and "non-linguistic" perceptions of affiliation to impact learners' attention to linguistic variation and preference for novel labels from contrasting sources.

To successfully learn a lexical contrast, the listener must be able to disproportionately rely on data from accurate and unambiguous speakers: those for whom the category is most successfully inferred from the data. By inferring or observing the speaker's qualities, the listener is able to place the speaker's intended representation in a social context.

If accentedness influences the willingness of a child to interact with or trust information provided by a speaker we must ask - how do children identify their own native accent, and what is the role of beliefs about informant quality in this cognitive process?

To describe the acquisition of referential lexical representations, or words, traditional models idealize “the input” received by the child to be generated by a relatively uninformative random process, and idealize that the goal of learning is a single valid linguistic structure represented by that input. Effectively, these models are constructed on an abstraction that the available speech information is comprised of equally reliable observations about the world, from equally reliable sources, which may be recruited equally in the learning task, to recover a single native language structure. (Pinker, 1979; Fried and Holyoak, 1984; De Boer and Kuhl, 2003; Chater and Manning, 2006; Xu and Tenenbaum, 2007; Norris and McQueen, 2008; Frank, Goodman, and Tenenbaum, 2009; McMurray, Aslin, and Toscano, 2009; Rasanen, 2012; Pajak, Bicknell, and Levy, 2013).

All probabilistic models, of course, are idealizations which simplify the systems they are created to study. Models therefore necessarily fail to correspond to the precise causal structure of the systems they represent. The idealization that variation in the perception of sources is not a significant factor in word learning is perhaps most useful for describing the development of subjects in an environment which supports the adoption of a single linguistic system. However, it is not useful for describing the achievement of competence with multiple distinct linguistic systems. Suppose that a listener is sensitive to non-linguistic, social classifications of linguistic sources - what will be the impact on their perception and acquisition of referential lexical representations?

In reality, speech communities, speech environments and the speakers who populate them are heterogenous and learners will encounter a variety of speakers, who may speak with differing sociolects, ethnolects, regional dialects or other linguistic variations some of which are not part of the listener's target dialect. Language learners ultimately achieve both productive and perceptual linguistic competence reflecting their own social identity, rather than acquiring a language which represents an average of all the linguistic variation they have observed. I wish to model first language acquisition in such a way as to account for how typically developing monolingual language learners judgements of speech acquire consistent patterns of not just *within* but *between* dialect judgements about the representative quality of speech.

In uniting the literature on epistemic trust and natural pedagogy with early phonetic and word learning, I will show how children might use Bayesian inference to form beliefs about the relative quality of linguistic informants, and how such judgements would necessarily affect the child's early acquisition of word forms. Assuming that adults do in fact encode the referential content of words without direct representation of information regarding speaker identity, I will nevertheless demonstrate that beliefs about variation in speaker credibility implied by potential speaker membership in social groups exert an influence on word learning.

2.2 A Dialectal Word Learning Problem

Suppose a language learner is faced with the following problem: one of several speakers has produced a word in the presence of an known object. The learner must determine what word was produced and whether it properly labels this object. For example, suppose the speakers' utterances are distributed among three distinct phonetic categories. The speaker may have intended to say "dog" $[d\text{ɑː}g]$, "dawg" $[d\text{ɔː}g]$, or "dug" $[d\text{ʌ}g]$.

Let us also suppose that these categories are confusable, so with some probability, each production of a given phoneme may yield an completely ambiguous phonetic form, i.e. the speaker will pronounce the word with an unintelligible, or masked vowel /$d - g$/.

To simplify the problem, we begin by supposing that the speech is known to be a labeling action with only one potential referent. Assuming for the moment that just *one* of the distinctive variants is valid in a given dialect, each speaker will produce tokens from only one of the speech categories 'dog" $/d\text{ɑː}g/$, "dawg" $/d\text{ɔː}g/$, or "dug" $/d\text{ʌ}g/$, but all speakers may also produce the ambiguous token /$d - g$/. The rate at which the ambiguous token occurs varies with the category, and is denoted by probability p_X, where X is an index on the set of unambiguous tokens.

If the learner is exposed to speakers of all three dialects, but does *not* differentially weight their inputs, we should expect the learner to acquire a distribution over unambiguous phonetic forms that is a mixture of all three distinctive vowels. We would also expect the learner to predict ambiguous phonetic forms at a rate which is intermediate between the frequencies at which these ambiguous forms are produced by each speaker. In other words, the frequency of exposure to each speaker type will govern the learned forms.

In effect, if all linguistic informants are equally valid, then it is the amount of speech provided by each informant, and not any quality of the informants sources which will predict judgements about which informants' dialects our learner acquires.

However, if the learner's goal is to successfully acquiring a single productive pattern used by a subset of the population while retaining the ability to perceive contrasts in alternate dialects, then they must determine *which* of the speakers are using this privileged distribution of labels, and preferentially attend to data from those speakers.

2.3 The object label inference problem in a social context

In the given example there are potentially multiple utterances labeling a known referent. In this situation, the listener correctly believes the speaker is referring to a dog, producing speech tokens using one of three possible distributions over utterances. However, it is unknown to the learner which is the "correct" pattern of pronunciation. We will investigate the impact of beliefs about speaker characteristics on inferences about labeling by focusing on two speaker characteristics which predict how often the speaker will produce forms that are correct and unambiguous - knowledgeability and representativeness.

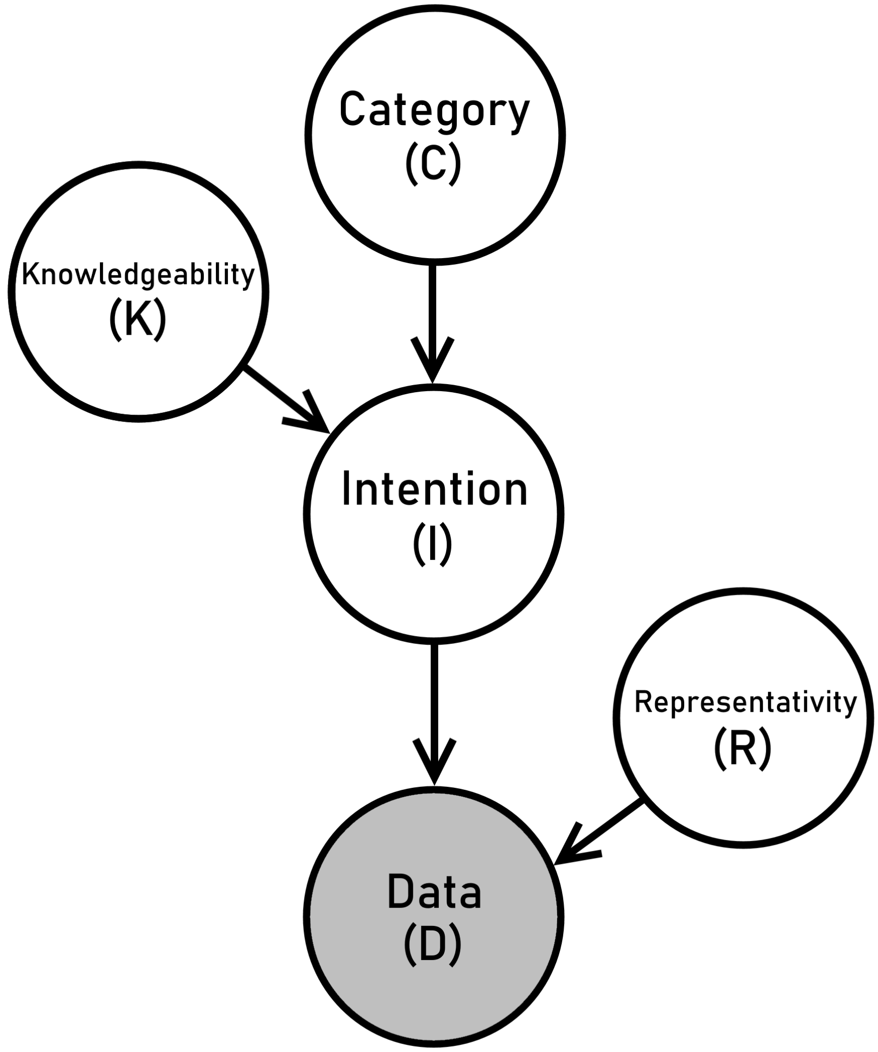
2.3.1 Speaker Characteristic: Knowledgeability

The first speaker characteristic, knowledgeability, predicts how often the speaker selects the proper intention, i.e. the one which matches the category. We will explore a simple case, assuming that there are only two levels of knowledgeability. *Knowledgeable* speakers possess the correct one to one mapping between intentions and categories. By contrast, *unknowledgeable* speakers possess a one to one mapping between intentions and categories which is randomly generated prior to the beginning of the experiment. This means that unknowledgeable speakers have some probability of behaving in a knowledgeable manner, dependent on the similarity of their mapping to that possessed by knowledgable speakers.

For each listening trial suppose an index variable x on the category set indicates the correct category, while a separate index y indicates the category intentionally represented by the speaker. The speaker may then either represent the correct category $(x = y)$, or an incorrect category $(x \neq y)$. The input will be generated by a mixture of speakers whose intentions are always accurate, and those whose intentions are often inaccurate. The probability that the speaker's intentions are accurate is determined by the variable of knowledgeability. A knowledgeable speaker $(K = 1)$ will always select the correct intention $P(x = y|K = 1) = 1$ whereas an unknowledgeable speaker $(K = 0)$ will select an intention at chance $P(x = y|K = 0) = \frac{1}{n}$. The probability of a speaker providing a correct or incorrect intention is therefore a function of k, where $P(K = 1) = k$.

Having selected one of the n possible intentions I with index y, The speaker then generates a speech token D with index z. Barring knowledge of the intention or category, we can describe this token as being drawn from a set of size $n + 1$ containing one unambiguous token for each possible intention, and an additional ambiguous token. Speakers may only produce unambiguous tokens which match their selected intention, narrowing this set to a size of 2: speakers probabilistically produce tokens which are either clearly indicative of the chosen category $(D_{z=y})$ or which are ambiguous between all possible intentions ($D_{z=0}$).

Figure 2.1: Graphical model



2.3.2 Speaker Characteristic: Representativeness

We will contrast speakers in a second dimension by specifying two levels of representativeness, a feature we use to describe how often the speaker produces ambiguous tokens. ”Unrepresentative” speakers will produce ambiguous tokens at a baseline rate. “Representative” speakers will produce data in a pedagogical fashion, decreasing the incidence of ambiguous tokens.

The input will be generated by a mixture of speakers who are representative and unrepresentative samplers. An unrepresentative speaker ($R = 0$) does not behave pedagogically, and will select the unambiguous token at some base rate which depends on the intention. $P(D_{Z=Y}|I_Y) = p_y$. Otherwise, they produce an ambiguous token $P(D_{Z=0}|I_Y) = (1 - p_y)$. Given that there are n possible categories and n possible intentions, there are n^2 ways for a coin and intention to be selected. However, there are only two possible observations for each intention. In each trial, the speaker either produces an ambiguous token, or an unambiguous token matching the selected intention. We define the representative and unrepresentative speakers' behavior according to the incidence of these 2 possible outcomes.

When referring to pedagogical samplers, we call them "representative," referencing the Bayesian definition of representativeness (Tenenbaum and Griffiths, 2001). For ease of reference, the non-pedagogical samplers are labeled "unrepresentative," although the data from "unrepresentative" speakers *is* measurably representative of the same distribution as those we have termed "representative" just comparatively weakly. It would be clumsy to refer to them as "less representative," so we will continue with the established terminology.

Recall, our listener's task is to avoid acquiring the pronunciations exhibited by unknowledgeable speakers and preferentially adopt the productive patterns exhibited by the most useful informants. This is modeled as an inference regarding the knowledgeability and representativity of the speaker given the observed productions, with each contribution weighted accordingly. Observations from speakers of different types must be adequately identified and weighted according to their informational value.

The "Unrepresentative" Speaker

As an example, imagine a set of three categories, all equally likely to occur, but with varying rates of ambiguity, as specified in 2.1 and 2.2.

$$c_1 = 1/3,\ c_2 = 1/3,\ c_3 = 1/3 \tag{2.1}$$

$$p_1 = 1/4,\ p_2 = 1/2,\ p_3 = 3/4 \tag{2.2}$$

For each sampled token, there are $2 \times n$ possible outcomes - the speaker may produce either an ambiguous or an unambiguous token with the intention of representing any of the n categories. The graph on the left below shows the rate at which each of these outcomes occurs for an unrepresentative speaker given an intention, simply matching the rates established in 2.2.

Figure 2.2 shows the incidence of ambiguous (D0) and unambiguous data (D1,D2,D3) for unrepresentative speakers given a selected intention. The three categories have different distributions of ambiguous data, but these proportions are constant across values of K. Effectively, both knowledgeable and unknowledgeable unrepresentative speakers have identical distributions of $(D_Z|I_Y)$.

The ”Representative” Speaker

Representative speakers are different from unrepresentative speakers in that the data they produce improves the rate at which listeners can successfully guess I from the distribution of data. The representative speaker gives data proportional to the posterior probability of observing that data, given the target hidden variables, as in equation 2.3, describing the posterior probability of believing in category C_X given an observation of coin face D_Z. Representative speakers behave pedagogically, producing fewer ambiguous tokens than unrepresentative speakers and increasing the listener’s posterior probability on the true values of the hidden variable C.

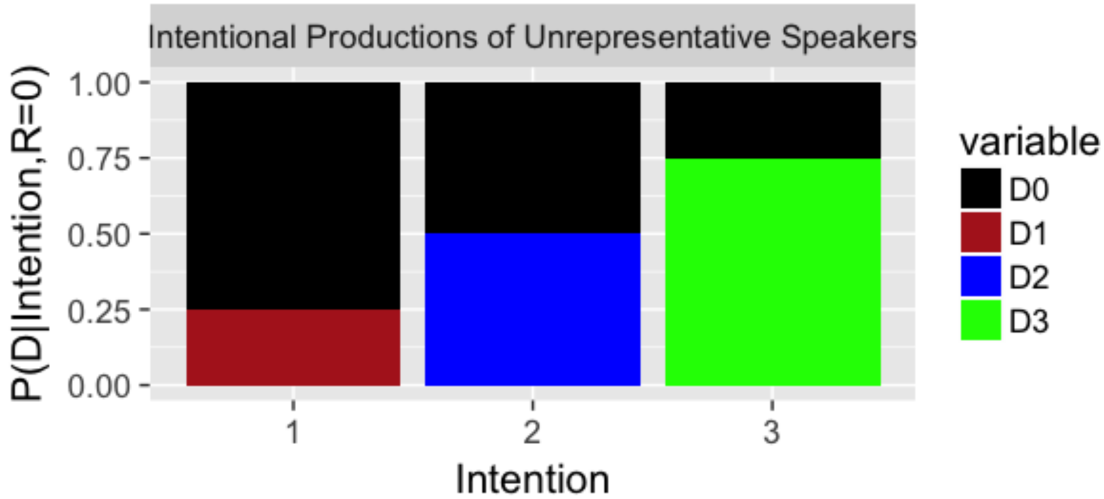


Figure 2.2: Distribution of data over speakers given intention, low representativity: $P(D_Z|R=0)$

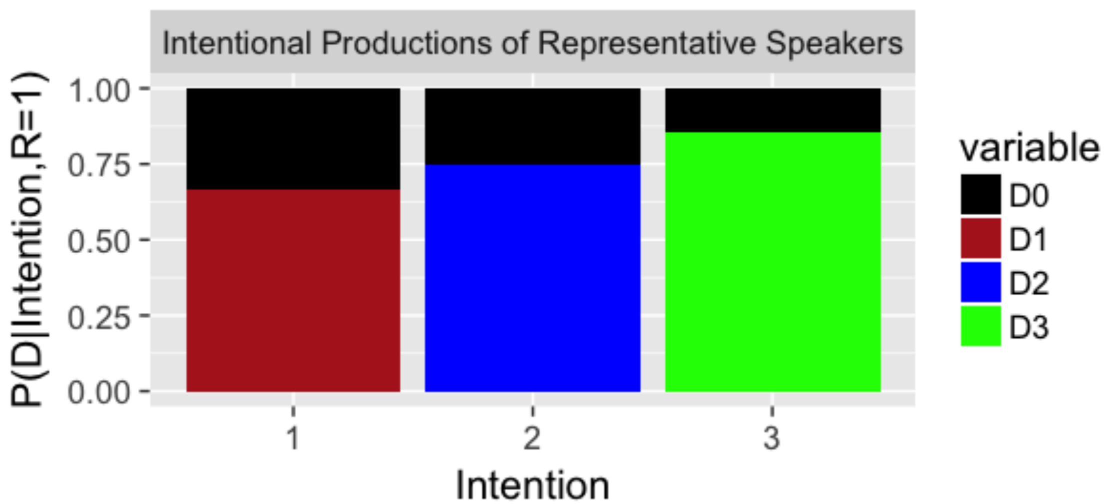


Figure 2.3: Distribution of data over speakers given intention, high representativity: $P(D_Z|R=1)$

The graph in figure 2.3 shows the incidence of ambiguous (D0) and unambiguous (D1,D2,D3) data given a selected intention for representative speakers.

$$P_{speaker}(C_X|D_Z,R_1) = \frac{\frac{P(D_Z|C_X)P(C_X)}{\sum_{C'} P(D_Z|C')P(C')}}{\sum_{D'} \frac{P(D'|C_X)P(C_X)}{\sum_{C'} P(D_Z|C')P(C')}} \tag{2.3}$$

The constant given in the denominator of 2.3 adjusts the posterior so that the relative probability of each hypothesis of a value for C is normalized with respect to the $(n+1)$ expressions of those values in D.

The proportions of ambiguous to unambiguous data are not equivalent for representative and unrepresentative speakers, differing systematically across categories. In effect, the representative speaker provides a skewed data set, emphasizing unambiguous data, licensing stronger inferences from the learner. A representative speaker $(R = 1)$ behaves pedagogically, selecting the unambiguous token at a higher rate than an unrepresentative one. For simplicity we will assign this probability to the variable q_X, where X indexes the intentions. Therefore the probability of a representative speaker producing an unambiguous token is given by $P_{speaker}(D_Z|R_1, I_X) = q_X$, and the probability of an ambiguous token is given by $P_{speaker}(D_0|R_1, I_X) = (1 - q_X)$.

2.4 An analogy: flipping coins

Now we can reframe our learning problem: our learner, observing a label, trivially determines the referent, but still must determine which word form was selected, *and* what sort of speaker likely made the selection, before judging how to weight the observed input.

The model described here will allow us to address the dual problems of how to identify and manage data from four different types of speakers - those who are both knowledgeable and representative, neither knowledgable nor representative, representative and unknowledgeable, or knowledgeable and unrepresentative. We begin with the broad assumption that the selection of the "correct" category in any trial is independent of informant qualities - that is to say, the listener's target dialect is independent of the characteristics of any randomly sampled speaker.

To create a simplified set-up of the word-learning problem, rather than describing phonetic categories as distributions over continuous acoustic values, our phonetic categories will be distributions over discrete variables. Suppose the listener, observing a dog, believes that there are three possible labels for this object, with the distinct phonemic forms "dog" $[d\text{ɑː}g]$, "dawg" $[d\text{ɔː}g]$, or "dug" $[d\text{ʌ}g]$. Each of these phonetic forms may also with some probability be produced in an ambiguous phonetic form $[d-g]$. We can imagine the speech categories as a collection of coins, where the heads correspond to unique distinct observable pronunciations. This is expressed with the variable D_X, where X indicates the X^{th} category. So a production of $[d\text{ɑː}g]$ corresponds to an observation of D_1, a production of $[d\text{ɔː}g]$ to D_2, etc. By contrast, the tail face of each coin is identical, and is denoted with D_0, the ambiguous token. The listener's dialect is described by a distribution over the coins and their observable faces, $P(C_X)$ and $P(D_X, D_0|C_X)$.

A trial proceeds thusly: The correct coin for the trial is specified by the dialect of the *listener*, sampling from the distribution of C_X. The listener's dialect then also determines the expected distribution of pronunciations $P(D|C_X)$. A listener with an uninformative prior may expect "dog" $[d\text{ɑː}g]$, "dawg" $[d\text{ɔː}g]$, or "dug" $[d\text{ʌ}g]$ to appear interchangeably, describing a uniform distribution for C_X.

The speaker must then select a coin. The speaker's coin selection is encoded with the variable (I) for intention. We index the categories and intentions allowing us to pair each intention, or speaker attempt, to a single matching category. The speaker is guaranteed to select the correct coin, (selecting intention I_X in the context of category C_X) with some probability $P(K = 1) = k$. With complementary probability, $P(K = 0) = 1 - k$ the speaker selects a random intention, which may or may not match C_X.

Once the speaker has selected an intention I, they articulate their pronunciation by tossing this coin, revealing the upward face (D), or observed pronunciation. A speaker who is correctly producing tokens with intention I_X will produce tokens D_X or D_0, whereas a speaker producing tokens from an incorrect intention $I_{\neg X}$ will either produce tokens $D_{\neg X}$ or D_0. The listener, upon observing a pronunciation D must infer the identity of category C, given that there is uncertainty about the characteristics of the speaker.

For unrepresentative speakers, each coin comes up heads according to the distribution of $(D_X|I_X)$ in 2.2. However, representative speakers sample helpfully, producing fewer ambiguous tokens. We can imagine representative speakers as using a modified set of coins, where each is weighted more heavily towards the heads side.

In this analogy, the learner observes the outcomes of coin tosses from one or more speakers, and must determine the identity of the correct coin. For each coin toss she observes, she must additionally infer whether it was supplied by a knowledgeable and/or representative informant.

In the next section we will outline this inference, and demonstrate that this model predicts that sensitivity to ostensive cues is gated by epistemic trust. A learner may only consider knowledgeable speakers to be adequate informants, so any assumption of pedagogical value must be predicated first on this capacity, and second upon the informant's relative representativity. Among knowledgeable speakers, increasing representativity is associated with increasing usefulness.

Chapter 3

Solving the Socio-phonetic Inference Problem

3.1 Solving the socio-phonetic inference problem

In order to model the perceptual categorization of our listener, we begin by supposing they observe some data D, and then must infer which is the correct category C. Using Bayes theorem, we define the posterior probability of the category given the data in equation 3.1.

$$P(C|D) = \frac{P(D|C)P(C)}{\sum_{C'} P(D|C')P(C')} \tag{3.1}$$

This model describes a listener who attends to the speech signal irrespective of information characterizing its source. To reflect a social dimension to this process, we begin by amending this model to reflect knowledge about distinctive behavior from different types of informants.

We suppose there are some number of groups which categorize the informants, and each group G has its own characteristic rates of data production which may be conditioned on C. To determine the probability of a given category, we may sum over the likelihoods that the data was produced by an informant from each group, weighting each with the prior probability of informant membership in that group. This sum is given in equation 3.2.

$$P(C|D) = \sum_{G} \frac{P(D|C,G)P(C)P(G)}{\sum_{C'G'} P(D|C',G')P(C')P(G')} \tag{3.2}$$

As outlined in section 2.4, the groups currently under consideration are speakers who are either knowledgeable, representative, both, or neither. The following table restates the conditional probability of unambiguous (D_x) and ambiguous (D_0) data, as it was given earlier.

Table 3.1: Probability of data given category and group membership

G	K	R	$P(G)$	$P(D_X \| C_X, G)$	$P(D_0 \| C_X, G)$
1	1	1	kr	q_x	$(1 - q_x)$
2	1	0	$k(1 - r)$	p_x	$(1 - p_x)$
3	0	1	$(1 - k)r$	$\frac{q_x}{n}$	$\sum_x \frac{(1-q_x)}{n}$
4	0	0	$(1 - k)(1 - r)$	$\frac{p_x}{n}$	$\sum_x \frac{(1-p_x)}{n}$

To update the inference problem, we now characterize our listener as attempting to infer both the correct category C, as well as which group G the informant belongs to. However, as the listener observes additional speakers, the hypothesis space grows exponentially, rendering the inference intractable.

If, for example, the listener observes ambiguous data from m different informants, each may be either knowledgeable, representative, or both. The result is $n(2 \times 2)^m$ hypotheses. As the number of informants increases linearly, the number of possible scenarios grows exponentially.

In order to solve this problem despite its computational complexity, we will implement a sampling algorithm to estimate the joint posterior. In the next section we will give this joint inference for a single observation. In this case, G is synonymous with combinations of K and R, so the joint posterior for category and group given the data is $P(C, K, R|D)$. Finally, we will show that both unrepresentative and representative informants who are unknowledgeable have equally low utility, allowing us to redefine the groups given in table 3.1 to reflect a hierarchy of utility.

Table 3.2: Probability of data given category and group membership: simplified

G	K	R	$P(G)$	$P(D_X \mid C_X, G)$	$P(D_0 \mid C_X, G)$
1	1	1	kr	q_x	$(1-q_x)$
2	1	0	$k(1-r)$	p_x	$(1-p_x)$
3	0	-	$(1-k)$	$\frac{q_x+p_x}{n}$	$\sum_x \frac{(1-q_x)+(1-p_x)}{n}$

3.2 Joint inference of category and informant type

3.2.1 For a single observation

Supposing a single observation, Bayes Theorem defines the joint posterior probability distribution of category and informant knowledgeability over $4n$ possible scenarios. For each combination of category and knowledgeability, the posterior probability of that scenario given the observation is weighted by the prior probability assigned to that combination, and divided by the normalizing constant that is the probability of the observation summed over all possible scenarios. This calculation is depicted in equation 3.3 below. The assumption of a uniform prior causes those terms to cancel, leaving us with a simplified form.

$$P(C_x, K_w, R_y \mid D_z) = \frac{P(D_z \mid C_x, K_w, R_y)}{\sum_{x',w'y'} P(D_z \mid C_{x'}, K_{w'}, R_{y'})} \tag{3.3}$$

In this example, with a single data point from a single speaker, the learner must infer a belief about the category (C), and beliefs about the qualities of the speakers, knowledgeability and representativity (K and R). A distribution of interpretations over possible triplets $(C_{x'}, K_{w'}, R_{y'})$ is given for each of the $n+1$ possible single observations.

3.2.2 Increasing the number of speakers

As we will show in section 3.5, in the presence of uncertainty about which data comes from knowledgeable speakers, there is effectively uncertainty about which speaker data can be safely ignored. Any pool of data gathered from multiple informants of unknown quality will be expected to potentially vary in information density. We may assume ordering does not matter as it does not effectively impact the posterior probability of C and K. Our model of the listener infers the knowledgeability of each speaker based on the dataset in its entirety. As a result it is not possible for the learner to entertain the belief that two speakers with unambiguously contradictory speech tokens are both knowledgeable.

When the number of speakers is increased beyond one, D_z becomes $(D_{\vec{z}})$, K_w becomes $K_{\vec{w}}$ and R_y becomes $R_{\vec{y}}$. The number of possible scenarios quickly grows intractable.

To overcome this obstacle, will consider two simpler inference problems. In the first, labeled the "Litmus Test," to reach a determination about whether the speaker is knowledgeable and/or representative, the listener relies on pre-existing knowledge about the category. Inferring the speaker's characteristics while assuming that the category is known corresponds to the predictions of an account of epistemic trust. In the second inference problem, or the "Trust Fall" the listener must infer what the category is, supposing the informant qualities of knowledgeability and representativeness are known. Inferring the category while assuming that the speaker's knowledgeability is known corresponds to the predictions of natural pedagogy. Following Shafto et al. (2012) we will implement Gibbs Sampling to sample from the joint posterior distribution of categories and informant qualities.

Recalling our coin flip analogy from section 2.4, we can describe the "correct" label for the object as specified by the experimenter. The listener observes some number of data points before inferring this label, and therefore the distribution of C_X.

Suppose that there is a bin of coins where each is inscribed with "dog" ([dɑːg]) on one face, and a second bin with "dawg" ([dɔːg]) so inscribed. Both types of coins have an identical opposite face, or "tail," corresponding to a token [$dog/dawg$], ambiguously a representation of the head on either type of coin. Each speaker, according to their knowledgeability, then selects a coin (intention to convey the corresponding label) from one of the bins. Knowledgeable speakers will always select the correct coin, whereas unknowledgeable speakers select it at chance. Each time that speaker is called upon to make a production, they toss their chosen coin, revealing the pronunciation observed by the listener. The listener, upon observing a pronunciation D must infer the identity of the correct label C, given that there is uncertainty about the characteristics of the speaker.

For unrepresentative speakers, each coin comes up heads according to the distribution of $(D_X|I_X)$ in 2.2. However, representative speakers sample helpfully, producing fewer ambiguous tokens. We can imagine representative speakers as using a similar set of coins, with each weighted more heavily towards the heads side.

3.3 Litmus Test

In this inference problem, the category and speech token (C, D) are observed, while the listener must infer the informant qualities knowledgeability and representativeness (K, R). Given that K and R each take two possible values, we may therefore describe four types of informants.

Given the speech token and a label presumed to be correct, the listener judges whether they are facing a knowledgeable and representative informant ($K = 1, R = 1$), a knowledgeable informant who is not representative ($K = 1, R = 0$), an informant who is unknowledgeable, but representative ($K = 0, R = 1$) or an informant who is neither knowledgeable, nor representative ($K = 0, R = 0$). The assumption that C is known provides us with a set of standards for our litmus test; for each pair of token and category, there is a probability distribution over all possible informant types.

Fixing the speaker qualities, K and R, we may investigate $n(n+1)$ scenarios at a time. The distribution is simple to calculate for unrepresentative speakers. Table 7.9 gives the unnormalized posterior probabilities of the observed token and category (D_Z, C_X), assuming $K=1$ and $R=0$. Table 3.4 gives the unnormalized posterior probabilities of these same pairs for informant qualities $K=0$ and $R=0$. Equation 3.4 gives a formula for normalizing these values.

$$P(K,R|C_X,D_Z) \propto \frac{P(K,R)\sum_Y P(D_Z|C_X,K,R)P(I_Y|C_X,K)}{\sum_{K',R'} P(K',R')\sum_Y P(D_Z|C_X,K',R')P(I_Y|C_X,K_1)} \tag{3.4}$$

Table 3.3: The knowledgeable and unrepresentative informant

Z	X	$P(K_1, R_0) \sum_Y (D_Z \vert C_X, K_1, R_0, I_Y) P(I_Y \vert C_X, K_1)$	$P(K_1, R_0 \vert C_X, D_Z)$
0	1	$k(1-r) \times ((1-p_1)(1) + (1-p_2)(0) + ... + (1-p_n)(0))$	$k(1-r) \times (1-p_1)$
0	2	$k(1-r) \times ((1-p_1)(0) + (1-p_2)(1) + ... + (1-p_n)(0))$	$k(1-r) \times (1-p_2)$
$\vdots$	$\vdots$	$\vdots$	
0	n	$k(1-r) \times ((1-p_1)(0) + (1-p_2)(0) + ... + (1-p_n)(1))$	$k(1-r) \times (1-p_n)$
1	1	$k(1-r) \times (p_1(1) + p_2(0) + ... + p_n(0))$	$k(1-r) \times p_1$
1	2	$k(1-r) \times (p_1(0) + p_2(0) + ... + p_n(0))$	0
$\vdots$	$\vdots$	$\vdots$	
1	n	$k(1-r) \times (p_1(0) + p_2(0) + ... + p_n(0))$	0
$\vdots$	$\vdots$		
n	n	$k(1-r) \times (p_1(0) + p_2(0) + ... + p_n(1))$	$k(1-r) \times p_n$

Knowledgeable informants never make overt errors - the only potentially misleading data they produce is ambiguous. For each category there are $n(n-1)$ combinations of intentions and categories which these informants never select and only n which they do. In this context, unambiguous tokens are strong evidence as to the speaker's knowledgeability.

Table 3.4: The unknowledgeable and unrepresentative informant

Z	X	$P(K_0, R_0) \sum_Y (D_Z \vert C_X, K_0, R_0, I_Y) P(I_Y \vert C_X, K_1)$	$P(K_0, R_0 \vert C_X, D_Z)$
0	1	$(1-k)(1-r) \times \frac{1}{n}((1-p_1) + (1-p_2) + ... + (1-p_n))$	$(1-k)(1-r) \times (1 - \frac{\sum_z p_z}{n})$
0	2	$(1-k)(1-r) \times \frac{1}{n}((1-p_1) + (1-p_2) + ... + (1-p_n))$	$(1-k)(1-r) \times (1 - \frac{\sum_z p_z}{n})$
$\vdots$	$\vdots$	$\vdots$	
0	n	$(1-k)(1-r) \times \frac{1}{n}((1-p_1) + (1-p_2) + ... + (1-p_n))$	$(1-k)(1-r) \times \frac{(n - \sum_z p_z)}{n}$
1	1	$(1-k)(1-r) \times (p_1 * \frac{1}{n} + 0 + ... + 0)$	$(1-k)(1-r) \times \frac{p_1}{n}$
1	2	$(1-k)(1-r) \times (0 + p_2 * \frac{1}{n} + ... + 0)$	$(1-k)(1-r) \times \frac{p_2}{n}$
$\vdots$	$\vdots$	$\vdots$	
1	n	$(1-k)(1-r) \times (p_1 * \frac{1}{n} + 0 + ... + 0)$	$(1-k)(1-r) \times \frac{p_1}{n}$
$\vdots$	$\vdots$		
n	n	$(1-k)(1-r) \times (0 + 0 + ... + p_n * \frac{1}{n})$	$(1-k)(1-r) \times \frac{p_n}{n}$

Unknowledgeable informants produce the correct intention at chance and are responsible for all unambiguously errorful data. The unknowledgeable informant effectively selects from all n^2 possible combinations of intention and category.

The knowledgeable and representative informant always selects the correct intention, but distributes tokens over categories differently, providing fewer ambiguous tokens, as illustrated in Figure 2.3.

Table 3.5: The knowledgeable and representative informant

Z	X	$P(K_1,R_1)\sum_Y (D_Z \mid C,K_1,R_1,I_Y)P(I_Y \mid C_X,K_1)$	$P(K_1,R_1 \mid D,C)$
0	1	$kr \times (\frac{\frac{(1-p_1)*c_1}{\sum_Y (1-p_Y)*c_Y}}{1+\frac{(1-p_1)*c_1}{\sum_Y (1-p_Y)*c_Y}})(1) + (0)(0) + ... + (0)(0))$	$kr * (1-q_1)$
0	2	$kr \times ((0)(0) + (\frac{\frac{(1-p_2)*c_2}{\sum_Y (1-p_Y)*c_Y}}{1+\frac{(1-p_2)*c_2}{\sum_Y (1-p_Y)*c_Y}})(1) + ... + (0)(0))$	$kr * (1-q_2)$
⋮	⋮	⋮	⋮
0	n	$kr \times ((0)(0) + (0)(0) + ... + (\frac{(1-p_n)*c_n}{(1-p_n)*c_n+\sum_Y (1-p_Y)*c_Y})(1))$	$kr * (1-q_n)$
1	1	$kr \times ((\frac{1}{1+\frac{(1-p_1)*c_1}{\sum_Y (1-p_Y)*c_Y}})(1) + (0)(0) + ... + (0)(0))$	$kr * q_1$
1	2	$kr \times ((0)(0) + (0)(0) + ... + (0)(0))$	0
⋮	⋮	⋮	⋮
1	n	$kr \times ((0)(0) + (0)(0) + ... + (0)(0))$	0
⋮	⋮	⋮	⋮
n	n	$kr \times ((0)(0) + (0)(0) + ... + (\frac{1}{1+\frac{(1-p_n)*c_n}{\sum_Y (1-p_Y)*c_Y}})(1)$	$kr * q_n$

Table 3.6: The unknowledgeable and representative informant

Z	X	$P(K_0, R_1) \sum_Y (D_Z \| C_X, K_0, R_1, I_Y) P(I_Y \| C_X, K_1)$	$P(K_0, R_1 \| D, C)$
0	1	$\frac{(1-k)r}{n}(\frac{\frac{(1-p_1)*c_1}{\sum_Y (1-p_Y)*c_Y}}{1+\frac{(1-p_1)*c_1}{\sum_Y (1-p_Y)*c_Y}} + ... + (\frac{\frac{(1-p_n)*c_n}{\sum_Y (1-p_Y)*c_Y}}{1+\frac{(1-p_n)*c_n}{\sum_Y (1-p_Y)*c_Y}})$	$(1-k)r * \frac{\sum_Y 1-q_Y}{n}$
$\vdots$	$\vdots$	$\vdots$	
0	n	$\frac{(1-k)r}{n}(\frac{(1-p_1)*c_1}{(1-p_1)*c_1+\sum_Y (1-p_Y)*c_Y} + ... + (\frac{(1-p_n)*c_n}{(1-p_n)*c_n+\sum_Y (1-p_Y)*c_Y})$	$(1-k)r * \frac{\sum_Y 1-q_Y}{n}$
1	1	$\frac{(1-k)r}{n}(\frac{1}{1+(\frac{(1-p_1)*c_1}{\sum_Y (1-p_Y)*c_Y}})$	$(1-k)r * \frac{q_1}{n}$
$\vdots$	$\vdots$	$\vdots$	
1	n	$\frac{(1-k)r}{n}(\frac{\sum_Y (1-p_Y)*c_Y}{(1-p_n)*c_n+\sum_Y (1-p_Y)*c_Y})$	$(1-k)r * \frac{q_1}{n}$
$\vdots$	$\vdots$	$\vdots$	
n	n	$\frac{(1-k)r}{n}(\frac{\sum_Y (1-p_Y)*c_Y}{(1-p_n)*c_n+\sum_Y (1-p_Y)*c_Y})$	$(1-k)r * \frac{q_n}{n}$

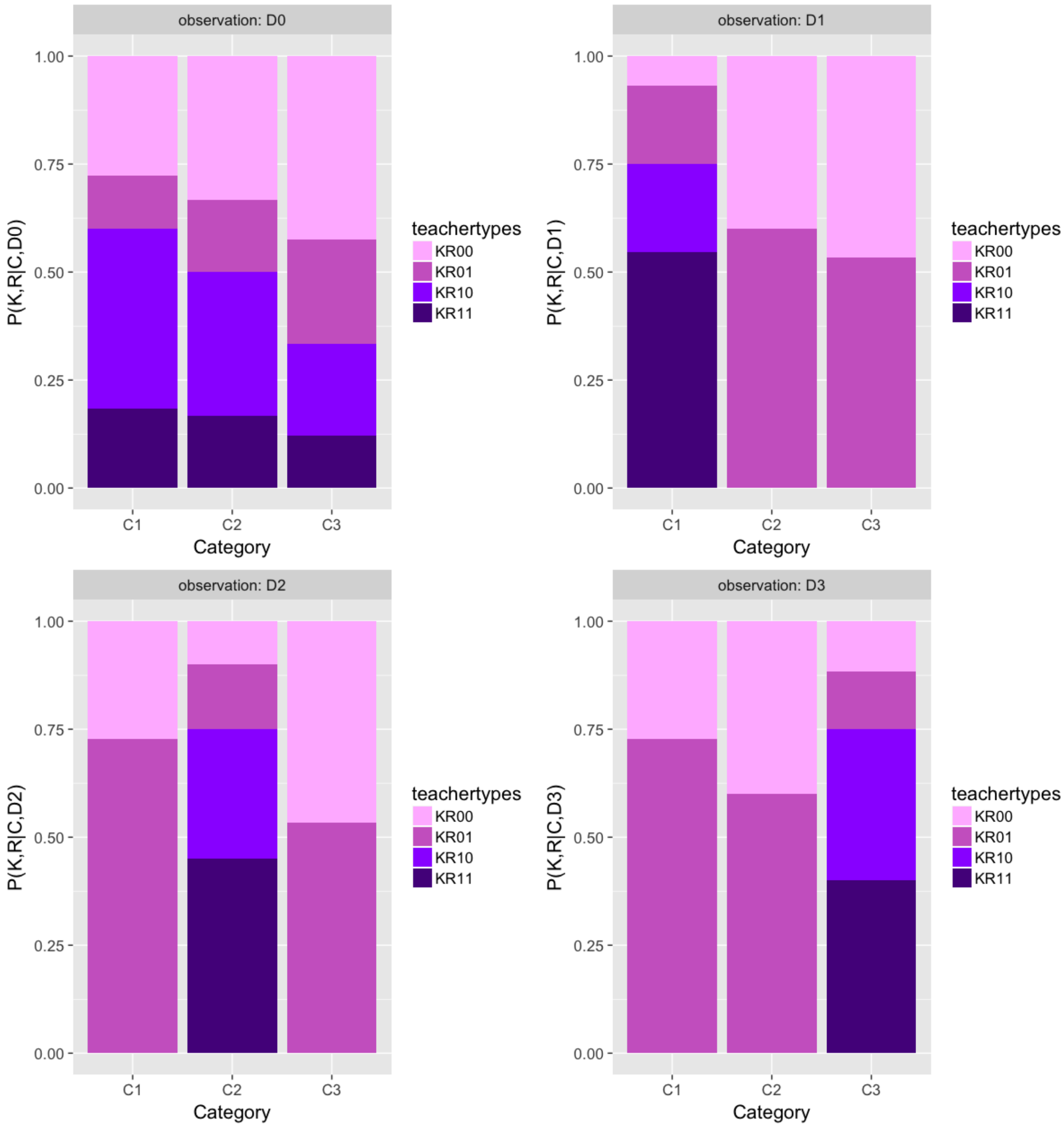
The following three simulations show what a learner with the priors given in equations 2.1 and 2.2 would believe about a speaker when presented with unambiguous tokens in the context of a known category. (i.e. the listener believes that “dog” $/d$ɑː$g/$ is the proper pronunciation, and then observes an utterance which is one of “dawg” $/d$ɔː$g/$, “dug” $/d$ʌ$g/$ or the ambiguous $/d-g/$). Each column in the graph represents the condition that the learner knows that category to be the correct one. The composition of each column shows the posterior probability of each teacher type after observing a particular pronunciation.

Each graph shows how the probability distribution across teachers changes when the listener interprets the data in the context of their belief about the category. The listener understands that knowledgeable speakers never produce tokens which are mismatched to the category, so in the context of believing that the first category is correct, observations of unambiguous tokens from categories two and three can only result in the belief that the informant is unknowledgeable, i.e. that the probability of either knowledgable informant type is zero. This is visible in the three graphs of 3.1 which illustrate listener analysis of unambiguous data - in each case, one of the three bars contains purple while the other two do not.

Each category is associated with a different rate of ambiguous productions. When observing an ambiguous production in the context of a given category, belief that the speaker is knowledgeable is positively correlated with the category's rate of ambiguous productions. In the context of believing that they have observed a category which is often realized as ambiguous, the listener will be more likely to attribute an ambiguous production to the speaker being knowledgeable, and in the context of believing they have observed a category which is often realized as unambiguous, the listener will be more likely to attribute an ambiguous production to the speaker being unknowledgeable. This is visible in the top left graph of figure 3.1; moving left to right, the purple bars constitute a decreasing proportion of the columns in accordance with the increasing likelihood of unambiguous data associated with each successive category.

Similarly, given the attribute of knowledgeability, the proportion of teachers who are assumed to be representative is related to the rate of unambiguous productions for the given category. As the rate of unambiguous productions given the category increases, supposing the speaker is unknowledgeable, the posterior probability that the speaker is representative increases. In each of the graphs showing the effects of observing unambiguous data, the size of the dark pink bar relative to the light pink bar decreases from left to right, along with the expected incidence of unambiguous data. Conversely, as the rate of ambiguous productions increases, supposing the speaker is knowledgeable, the posterior probability that that the speaker is representative shrinks. This contrast is illustrated by the decreasing proportion of dark purple to light purple and increasing proportion of dark pink to light pink in each of the graphs. This pattern is in accordance with the behavior of representative speakers producing relatively more unambiguous productions.

Figure 3.1: Probability of informant qualities given label $P(K, R|C, D)$



The simulation depicted in the upper left corner shows what a learner with an existing belief about the category would believe about a speaker when presented with an ambiguous token. The categories are ordered in decreasing likelihood of producing ambiguous data, so accordingly the relative likelihood of knowledgeability (indicated by the purple bars) given an ambiguous observation decreases as the expectation of unambiguous data increases.

3.4 Trust Fall

In this inference problem, the listener uses the model to infer the category (C), supposing the informant qualities, knowledgeability and representativeness are observed, in addition to the speech token (K, R, D). In other words, the learner observes the outcome of a coin toss, observes the speaker's qualities, and then must guess which category was named in the instruction to the speaker.

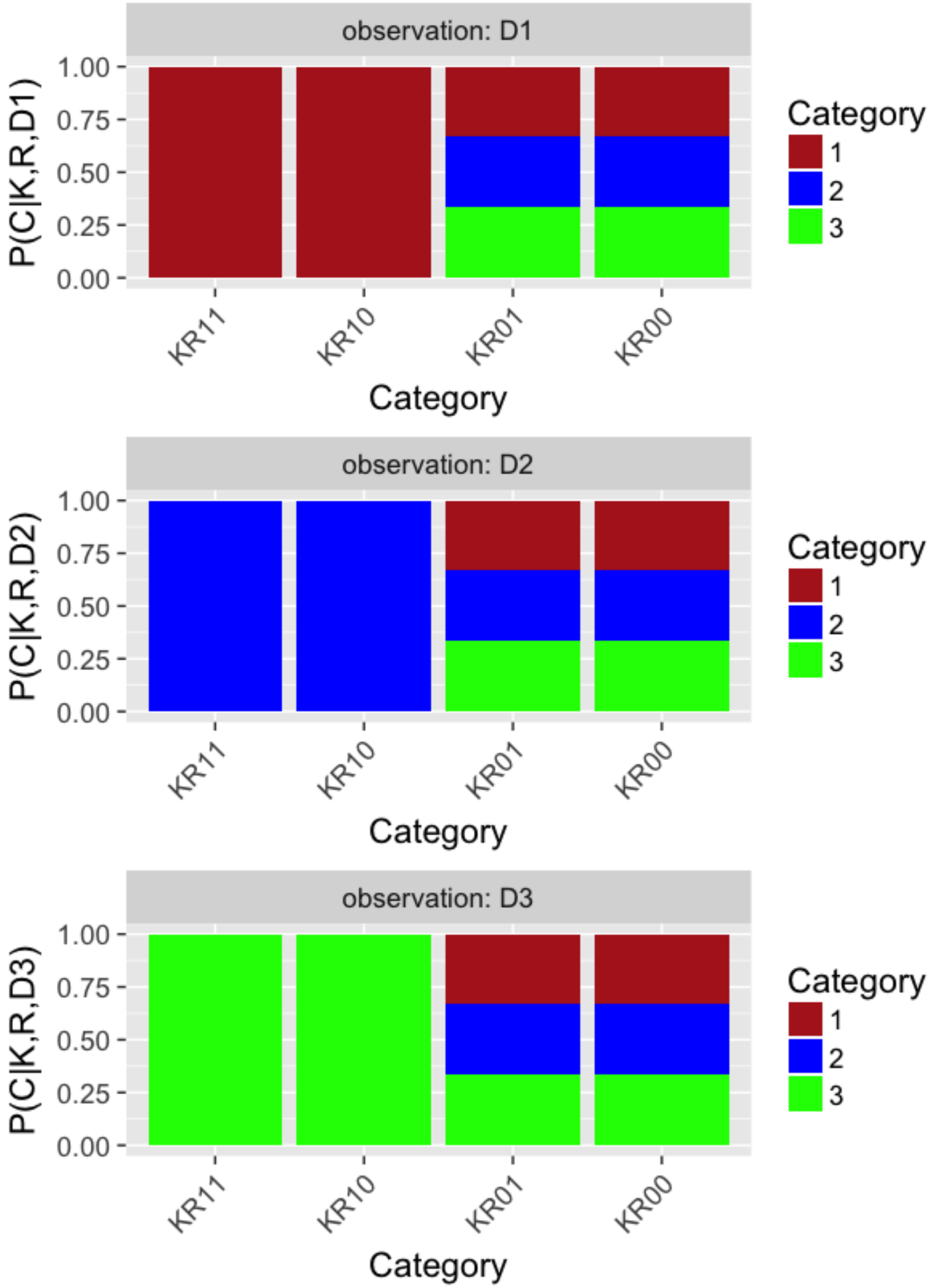
Tables 3.7 and 3.8 give the unnormalized posterior probability of each category given each of the four informant types.

Table 3.7: Trust fall with unambiguous data

X	$P(D_C \mid K_1, R_1, C_X)P(C_X)$		
1	$c_1 * \sum_X P(D_1 \mid K_1, R_1, C_1, I_X)P(I_X \mid K_1, C_1)$	$\frac{c_1 \sum_X (1-p_X)*c_X}{(1-p_1)*c_1+\sum_X (1-p_X)*c_X}$	$c_1 * q_1$
$\vdots$	$\vdots$	$\vdots$	
n	$c_n * \sum_X P(D_1 \mid K_1, R_1, C_n, I_X)P(I_X \mid K_1, C_n)$	$\frac{c_n \sum_{X'} (1-p_{X'})*c_{X'}}{(1-p_n)*c_n+\sum_X (1-p_X)*c_X}$	$c_n * q_n$
	$P(D_C \mid K_1, R_0, C_X)P(C_X)$		
1	$c_1 * \sum_X P(D_1 \mid K_1, R_0, C_n, I_X)P(I_X \mid K_1, C_1)$	$p_1 * c_1$	$c_1 * p_1$
$\vdots$	$\vdots$	$\vdots$	
n	$c_n * \sum_X P(D_n \mid K_1, R_0, C_n, I_X)P(I_X \mid K_1, C_n)$	$p_n * c_n$	$c_n * p_n$
	$P(D_C \mid K_0, R_1, C_X)P(C_X)$		
1	$c_1 * \sum_X P(D_1 \mid K_0, R_1, C_1, I_X)P(I_X \mid K_0, C_1)$	$\frac{c_1}{n} * \sum_{X'} \frac{\sum_X (1-p_X)*c_X}{(1-p_{X'})*c_{X'}+\sum_X (1-p_X)*c_X}$	$c_1 * \frac{\sum_X q_X}{n}$
$\vdots$	$\vdots$	$\vdots$	
n	$c_n * \sum_X P(D_n \mid K_0, R_1, C_n, I_X)P(I_X \mid K_0, C_n)$	$\frac{c_n}{n} * \sum_{X'} \frac{\sum_X (1-p_X)*c_X}{(1-p_{X'})*c_{X'}+\sum_X (1-p_X)*c_X}$	$c_n * \frac{\sum_X q_X}{n}$
	$P(D_C \mid K_0, R_0, C_X)P(C_X)$		
1	$c_1 * \sum_X P(D_1 \mid K_0, R_0, C_1, I_X)P(I_X \mid K_0, C_1)$	$\frac{c_1 * \sum_X p_X}{n}$	$c_1 * \frac{\sum_X p_X}{n}$
$\vdots$	$\vdots$	$\vdots$	
n	$c_n * \sum_X P(D_n \mid K_0, R_0, C_n, I_X)P(I_X \mid K_0, C_n)$	$\frac{c_n * \sum_X p_X}{n}$	$c_n * \frac{\sum_X p_X}{n}$

The following three simulations show how a learner's beliefs about the category are influenced by knowledge about the informant features, in the context of observing an unambiguous pronunciation. In each, an unambiguous pronunciation indicating a specific category has been observed. If the learner believes that their informant is knowledgeable, then they conclude that the category must be the indicated one, assigning a single hypothesis the posterior probability of 1, and all others to 0. If the informant is believed to be unknowledgeable, then the posterior probability for each category is identical to its prior. The quality of representativeness has no effect on the posterior probability of each category.

Figure 3.2: Probability of category given informant type, observation $P(C_X|K, R, D_Z)$



This fourth simulation, demonstrates how a learner's beliefs about the category are influenced by beliefs about the informant when the observation is of an ambiguous token. Table 3.8 shows the unnormalized posterior probability of each category given, an informant of each knowledgeability and representativeness. Figure 3.2 shows these values, normalized. In the first two columns, the knowledgeability of the speaker ensures that the distribution of categories deviates from the prior. In the last two columns, when the ambiguous token comes from an unknowledgeable speaker, the posterior probability of each category is identical to its prior.

$$P(K, R|D_0, C) = \frac{P(D_0|K', R', C)P(C)}{\sum_{C'} P(D_0|K', R', C')P(C')} \tag{3.5}$$

Table 3.8: Trust fall with ambiguous data

C	$P(D_0 \mid K_1, R_1, C)P(C)$		
1	$c_1 * \sum_X P(D_0 \mid K_1, R_1, C_1, I_X)P(I_X \mid K_1, C_1)$	$\frac{(1-p_1)*c_1^2}{(1-p_1)*c_1+\sum_{X'}(1-p_{X'})*c_{X'}}$	$c_1 * (1 - q_1)$
$\vdots$	$\vdots$	$\vdots$	
n	$c_n * \sum_X P(D_0 \mid K_1, R_1, C_n, I_X)P(I_X \mid K_1, C_n)$	$\frac{(1-p_n)*c_n^2}{(1-p_n)*c_n+\sum_{X'}(1-p_{X'})*c_{X'}}$	$c_n * (1 - q_n)$
	$P(D_0 \mid K_1, R_0, C)P(C)$		
1	$c_1 * \sum_X P(D_0 \mid K_1, R_1, C_n, I_X)P(I_X \mid K_1, C_1)$	$(1 - p_1)(c_1)$	$c_1 * (1 - p_1)$
$\vdots$	$\vdots$	$\vdots$	
n	$c_n * \sum_X P(D_0 \mid K_1, R_1, C_n, I_X)P(I_X \mid K_1, C_n)$	$(1 - p_n)(c_n)$	$c_n * (1 - p_n)$
	$P(D_0 \mid K_0, R_1, C)P(C)$		
1	$c_1 * \sum_X P(D_0 \mid K_0, R_1, I_X)P(I_X \mid K_0, C)$	$\frac{c_1}{n} \sum_X \frac{(1-p_X)*c_X}{(1-p_X)*c_X+\sum_{X'}(1-p_{X'})*c_{X'}}$	$c_1 * \frac{\sum_X 1-q_X}{n}$
$\vdots$	$\vdots$	$\vdots$	
n	$c_n * \sum_X P(D_0 \mid K_0, R_1, I_X)P(I_X \mid K_0, C)$	$\frac{c_n}{n} \sum_X \frac{(1-p_X)*c_X}{(1-p_X)*c_X+\sum_{X'}(1-p_{X'})*c_{X'}}$	$c_n * \frac{\sum_X 1-q_X}{n}$
	$P(D_0 \mid K_0, R_0, C)P(C)$		
1	$c_1 * \sum_X P(D_0 \mid K_0, R_0, I_X)P(I_X \mid K_0, C)$	$\frac{c_1}{n} * \sum_X (1 - p_X)$	$c_1 * \frac{\sum_X 1-p_X}{n}$
$\vdots$	$\vdots$	$\vdots$	
n	$c_n * \sum_X P(D_0 \mid K_0, R_0, I_X)P(I_X \mid K_0, C)$	$\frac{c_n}{n} * \sum_X (1 - p_X)$	$c_n * \frac{\sum_X 1-p_X}{n}$

3.5 Comparative value of data from different types of informants

To compare the benefit to the learner of data from each of the four types of informants, assuming that the we use a measure of information gain: the mutual information, measured in bits, between the category and the category given the data, given the informant attributes of knowledgeability and representativity.

Supposing that the informant has knowledgeability w and representativity y, the mutual information between the category and the category given the data is given in equation 3.6.

$$I(C; D|K_w, R_y) = H(C|K_w, R_y) - H(C|D, K_w, R_y) \tag{3.6}$$

The first term on the right hand side of equation 3.6 describes how much uncertainty the listener has about the category, given a set of informant attributes. This term will be identical for any pair of attributes K_w and R_y, owing to the fact that the category (C) is independent of both knowledgeability (K) and representativity (R).

$$H(C|K_w, R_y) = H(C) = \sum_i P(C_i) * I(C_i) \tag{3.7}$$

$$H(C|K_w, R_y) = H(C) = \sum_i P(C_i) * log\frac{1}{P(C_i)} \tag{3.8}$$

$$H(C|K_w, R_y) = \sum_i c_i * log(\frac{1}{c_i}) \tag{3.9}$$

Therefore the entropy of the category given any particular informant is identical to the entropy of the category. The informant's attributes alone do not reveal any information regarding which category has been selected. Substituting the value on the right hand side of 3.9 into equation 3.6, we are given 3.10.

$$I(C; D|K_w, R_y) = \sum_i c_i * log\frac{1}{c_i} - H(C|D, K_w, R_y) \tag{3.10}$$

The second term on the right hand side of equation 3.10, describes the entropy of the category given a single observation of data (D) from an informant with attributes K_w, R_y. This sum is presented in equation 3.11. The expected self-information of the category (C) is distributed over possible observations of the variable D, given K_w and R_y. In equation 3.12 this relationship is restated as the sum of the expected surprisal given the ambiguous production (tails) and the expected surprisal in each unambiguous production (heads).

$$H(C|D, K_w, R_y) = \sum_{d_z} P(d_z|K_w, R_y) * H(C|d_z, K_w, R_y) \tag{3.11}$$

$$H(C|D, K_w, R_y) = P(d_0|K_w, R_y)H(C|d_0, K_w, R_y) + \sum_{z>0} P(d_z|K_w, R_y)H(C|d_z, K_w, R_y) \tag{3.12}$$

Focusing first on knowledgeable informants, we will define the portion of entropy in the category contributed by ambiguous productions: $H(C|D_0, K_1, R_y)$. This corresponds to the first term on the right hand side of equation 3.12. The entropy in the category given an ambiguous production from a knowledgeable informant is described by the sum of the surprisal of each category value, weighted by the probability of the ambiguous production occurring for each category, as illustrated in equation 3.13.

$$P(d_0|K_1, R_y)H(C|d_0, K_1, R_y) = P(d_0|K_1, R_y)\sum_i P(C_i|d_0, K_1, R_y)log\frac{1}{P(C_i|d_0, K_1, R_y)} \tag{3.13}$$

Again, focusing on knowledgeable informants, we will define the portion of entropy in the category contributed by unambiguous productions: $\sum_{z>0} H(C|D_z, K_1, R_y)$. This corresponds to the second term on the right hand side of equation 3.12. Productions from knowledgeable speakers only produce uncertainty in the listener when those productions are ambiguous. Unambiguous tokens from knowledgeable speakers are always indicative of the correct category, as reflected by the zero valued term in 3.14.

$$\sum_{z>0} P(d_z|K_1, R_y)H(C|d_z, K_1, R_y) = \sum_{z>0} P(d_z|K_1, R_y) * 0 = 0 \tag{3.14}$$

Equations 3.15 and 3.16 describe the listener's entropy after observing data from unrepresentative and representative knowledgeable speakers, respectively.

$$H(C|D, K_1, R_0) = (\sum_i c_i * (1 - p_i)) \sum_j (1 - p_j) log \frac{1}{(1 - p_j)} + 0 \tag{3.15}$$

$$H(C|D, K_1, R_1) = (\sum_i c_i * (1 - q_i)) \sum_j (1 - q_j) log \frac{1}{(1 - q_j)} + 0 \tag{3.16}$$

Turning to unknowledgeable informants, the total expected uncertainty in a particular value of the category variable C_x given such an informant, is described by the uncertainty given one of $n + 1$ productions - an ambiguous production (d_0) or any unambiguous production (d_z).

$$H(C|D, K_0, R_y) = P(d_0|K_0, R_y) H(C|d_0, K_0, R_y) + \sum_{z>0} P(d_z|K_0, R_y) H(C|d_z, K_0, R_y) \tag{3.17}$$

The entropy in the category given an unknowledgeable speaker is identical to the entropy in the category, as a randomly selected unknowledgeable speaker effectively selects an intention at random.

$$H(C|D, K_0, R_y) = H(C) \tag{3.18}$$

Table 3.9 describes the expected information gain for a single observation from a speaker with each possible combination of attributes. Knowledgeable speakers who are also representative provide the most information, followed by knowledgeable speakers who are unrepresentative. Unknowledgeable speakers effectively provide no meaningful information, so their messages have an information content of 0. The rightmost column gives the expected information gain in bits, assuming the uniform distribution over categories and the specified distribution of unambiguous productions for each category specified in equations 2.1 and 2.2 in section 2.3.2.

Table 3.9: Mutual information in C, D given K, R

K_w	R_y	$I(C;D\|K_w,R_y) = H(C) - H(C\|D,K_w,R_y)$	in bits
1	1	$\sum_i c_i * log(\frac{1}{c_i}) - (\sum_i c_i * (1-q_i)) \sum_j (1-q_j) log \frac{1}{(1-q_j)}$	0.89
1	0	$\sum_i c_i * log(\frac{1}{c_i}) - (\sum_i c_i * (1-p_i)) \sum_j (1-p_j) log \frac{1}{(1-p_j)}$	0.53
0	1	$\sum_i c_i * log(\frac{1}{c_i}) - \sum_i c_i * log(\frac{1}{c_i})$	0
0	0	$\sum_i c_i * log(\frac{1}{c_i}) - \sum_i c_i * log(\frac{1}{c_i})$	0

Table 3.9 shows that the amount of information about the identity of the category expressed in a single data point depends on the quality of the informant. Knowledgeable informants always provide some information about the state of the category variable, while unknowledgeable informants never do. Therefore, representativity only indicates increased value for the listener when it is paired with knowledgeability. If the learner wishes to maximize efficiency at guessing the state of the category, then knowledgeable and representative informants will be preferred to knowledgeable and unrepresentative informants, with unknowledgeable informants all being equally dispreferred.

Data collected from a group of unknowledgeable speakers will be no more informative to the listener than simple guessing. The posterior probability distribution on the categories after viewing the data from unknowledgeable speakers will be identical to the prior. This data contributed by unknowledgeable informants contains an expected 0 bits of mutual information with the category. Therefore, without loss of generality, we may suppose listeners only consider data from speakers who are believed to be knowledgeable.

The ability to identify reliable informants is more or less synonymous with the ability to identify the production of reliable data. Supposing that reliable informants are the primary source of reliable data, the successful discrimination of reliable data without respect to source will nonetheless imply an ability to discriminate these informants from others who contrastively provide more unreliable data. Attending to the relative epistemic value of informants is therefore predicted to facilitate metacognitive biases, e.g. not only do listeners perceive speech and its reliability, they may also have a perception of the reliability of that very perception.

We established a framework in this chapter for modeling a socially grounded word learning problem. In the next chapter, I will use this framework to reinterpret a well-known finding from the infant speech perception literature, and show that this finding can actually be interpreted as providing evidence for social inference.

Chapter 4

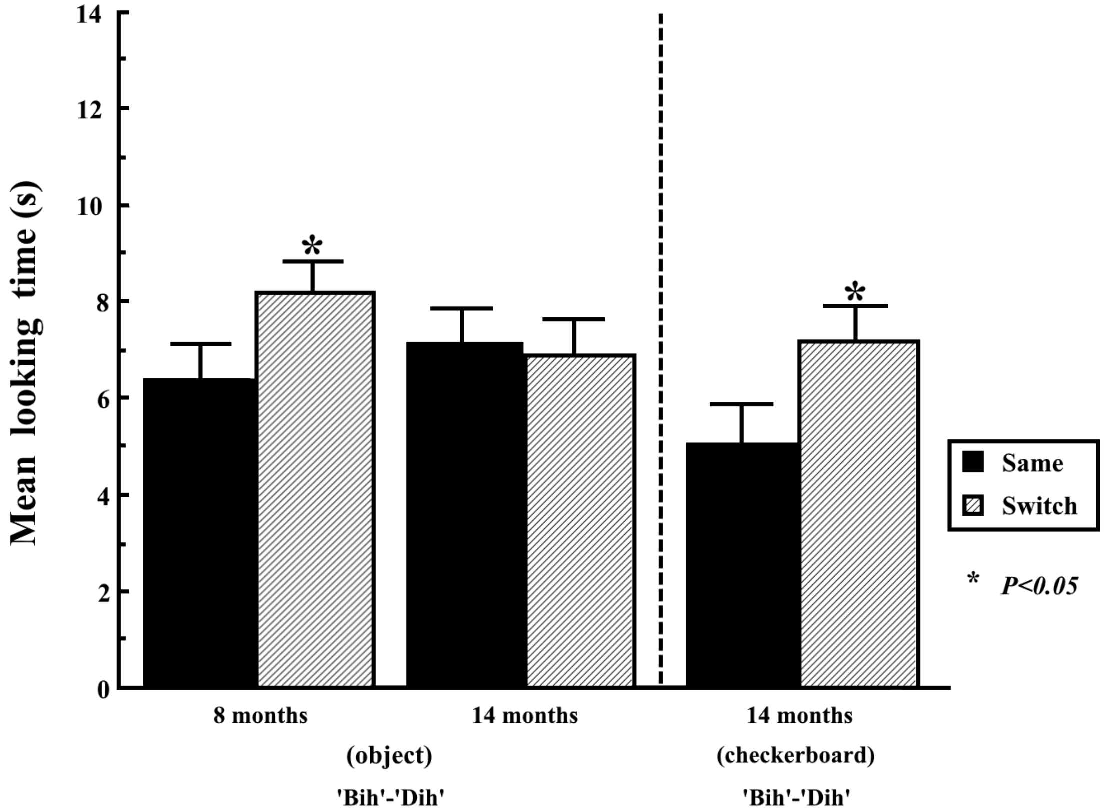
Simulating an Early Lexical Learning Task

4.1 Investigating early lexical representations

Typically developing infants have remarkable phonetic perception - just days after birth they show an adult-like pattern of categorical perception on many phonetic features, such as the voicing contrast in the minimal pair “ba” and “pa” (Eimas, Siqueland, Jusczyk, and Vigorito, 1971, 3968). By the end of the first year of life, infants phonetic perception abilities appear to become more specialized and native language-specific (J. F. Werker and Tees, 1984; P. K. Kuhl et al., 2006). By this age, they also have the ability to segment words from a fluent speech stream (Jusczyk and Aslin, 1995), and consistently prefer to listen to familiar words over novel ones (Halle and de Boysson-Bardies, 1994). Over the next year, the size of their vocabulary will begin to increase dramatically, along with the speed and accuracy with which they recognize words (A. Fernald, Pinto, Swingley, Weinbergy, and McRoberts, 1998).

But how is infant knowledge of native language sound patterns reflected in early lexical representations? A large body of work using an audiovisual habituation experiment called the Switch Task shows that infants who can perform well on a task discriminating two lexical neighbors, or words which differ by a single phoneme (i.e. "buk" and "puk") nevertheless do not consistently discriminate those same labels after being habituated to the presentation of these speech tokens as the labels of different objects. This difficulty does not appear for pairs of words which differ by multiple phonemes (e.g. "lif" and "neem"). Slightly older infants show significantly improved performance, with 17 month-old infants being successful at learning the phonetically similar words (J. F. Werker, Fennell, Corcoran, and Stager, 2002).

Figure 4.1: Stager and Werker (1997)



This pattern of findings appears to provide a window into the nature of infants' nascent lexical representations, and how they differ from those of more mature language users. The results were originally interpreted as strong support for the theory that infant speech representations undergo a functional reorganization. As the infant matures, their lexical representations were hypothesized to differ in character from the representations of phonetic detail they had previously relied on (Stager and Werker, 1997).

Many variations of the Switch task have been implemented. In the next few sections we will describe the Switch task in more detail, and focus our attention on a variant executed by Rost and McMurray (2009). I argue that these results are compatible with a new hypothesis - that the pattern of results obtained on the Switch task is evidence of infants performing a social inference. This hypothesis predicts that the failure to attend to the phonetic contrast is the result of selective inattention arising from the infant's belief that the informant is unlikely to be knowledgeable. We will call this the source-tracking hypothesis.

To provide support for our hypothesis, we will apply the model adapted from Shafto et al. (2012), which was described in chapter 3.1, to simulate the experiments in Rost and McMurray (2009). Lastly, we will discuss the source-tracking hypothesis in light of other variations on the Switch task, and implications for the developmental trajectory of the lexicon.

4.2 The Switch Task

In the audiovisual habituation procedure known as the Switch task, infants are presented with a repeated word paired with a visual display of a novel object. The presentation of stimuli continues until looking time drops below a preset level, meeting the habituation criteria. Then the infants' ability to discriminate the presented words is assessed using two types of trials:

On *same* trials, the subjects are exposed to the same object-word pairing(s) seen during habituation. On *switch* trials they again see one of the familiarized objects, but this time it is paired with a mismatched label.

If infants have successfully learned the object-label pairings, they are expected to dishabituate during switch trials, with longer looks to the presented object demonstrating that they notice and are surprised by the mismatch between label and object. Repeated reproductions of the Switch task have demonstrated that 14-month-olds are apparently able to learn pairs of labels sufficiently to dishabituate during switch trials when those labels differ by multiple phonemes, (i.e. ”lif” and ”neem”) but fail when the labels are lexical neighbors (i.e. ”buk” and ”puk”) (J. F. Werker and Tees, 1984; Stager and Werker, 1997; Rost and McMurray, 2009; Yoshida, Fennell, Swingley, and Werker, 2009). Figure 4.2 shows that the looking times of 14-month old infants to the object in same and switch trials are not significantly different. This result is surprising, considering that infants at this age have the ability to successfully attend to phonemic contrasts when the information is presented in a purely auditory context, without a visual object.

Stager and Werker argued their results suggested that as infants begin to learn words, the amount of phonetic detail infants recruit for speech perception tasks changes. At 14 months, the infants were apparently attending to the visual displays, yet demonstrating decreased sensitivity to phonetic detail. The authors hypothesized that as the infants begin mapping sounds onto meanings, they rely on more abstract representations to learn words. However, various modifications to the experimental procedure demonstrate that infants are capable of bringing finer perception of phonetic detail to bear in this task. For now I will focus on one such experiment, showing that exposure to multiple speakers during habituation supports 14 month old infants' success on the Switch task.

4.3 The effect of multiple speakers

Rost & McMurray demonstrated that the effect found in the Switch Task could be eliminated when subjects were exposed to exemplars from multiple voices during habituation. Supposing that infant phonological categories are still developing at this age, they posited that a more diverse data set, despite its complexity, would better facilitate the categorical learning. They trained infants on two lexical neighbors ("buk," "puk") in a Switch task, but with stimuli recorded from 18 speakers instead of only one. Unlike the 14-month olds who heard exemplars recorded in a single voice, infants in the condition with multiple speakers successfully discriminated lexical neighbors on the switch trials. Figure 4.3 shows that the looking time to same trials is significantly less for switch trials where the infants were exposed to multiple talkers.

Figure 4.2: Switch task, single speaker condition (Rost and McMurray, 2009)

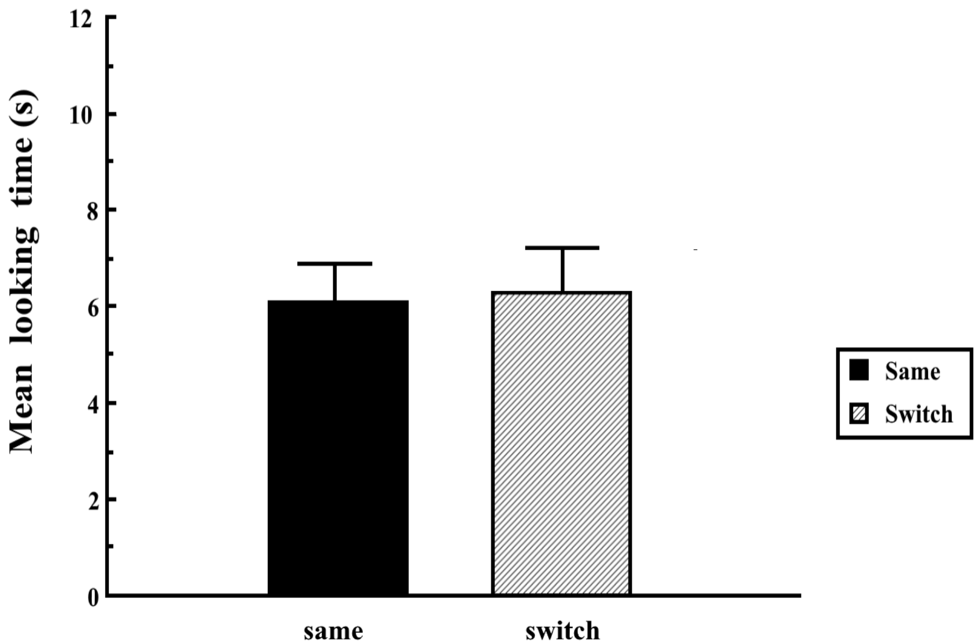
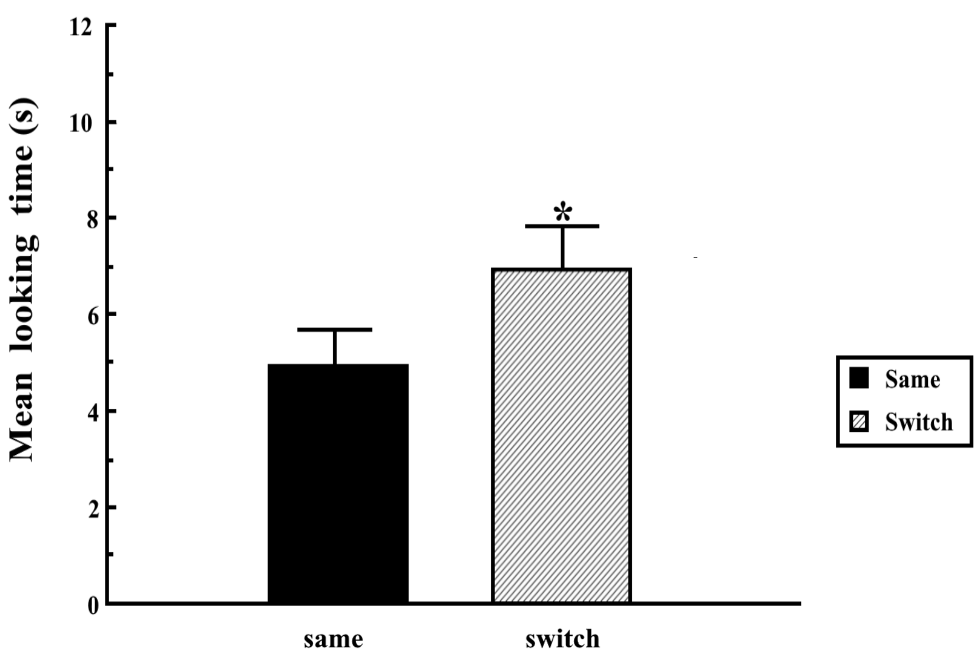


Figure 4.3: Switch task, multiple speaker condition (Rost and McMurray, 2009)



The authors hypothesized that the difficulty 14-month old infants show on the Switch Task is owed to a greater reliance on bottom-up processing of phonetic detail. They attributed the infants' success in the multiple-speaker condition to a greater availability of useful phonetic variation in the input.

Apfelbaum and McMurray (2011) subsequently modeled these results, demonstrating that they can be accounted for using basic associative learning principles. Cues which tend to have common values across all of a speaker's productions (i.e. cues to the speaker's identity) will be attested across both of the presented words. As a result, infants in the single speaker condition may therefore inappropriately attend to these cues, forming equally strong associations between them and each of the words.

Importantly, the Switch task is intended to evaluate whether the infants have learned a pair of words robustly enough to be surprised by a misnaming. To be successful on the task, the infant must react with increased looking time to these misnamings, or *switch trials* (i.e. the object previously labeled with "buk" is presented with "puk"). In an associative learning model, when the test token shares attributes with the habituation stimuli, these noncontrastive cues can cause partial activation of both categories and prevent the infant from recognizing the trial as a misnaming (Apfelbaum and McMurray, 2011). In effect, instead of their perceptual learning attending to the acoustic cues which indicate the speech contrast being tested, this account predicts that infants are attending to acoustic cues which indicate the identity of the speaker. In the multiple speaker condition, where cues to speaker identity are different with each observed token, it is not possible for the learner to make this mistake.

In this account, the infants are assumed to attend equally to both contrastive and non-contrastive cues from any given speaker. However, this explanation does not account for possible preferences infants may have for some informants over others. I show here that the results might also be explained by supposing the infants selectively attends to some speakers over others. Such preferences would predict the infants' selective *inattention* to both contrastive and noncontrastive cues. I will use the model outlined in section (3.1) to demonstrate how Rost & McMurray's results can be explained as the result of a social inference about the quality of linguistic informants.

In acquiring a specific dialect, not all sources will be equally useful to a language learner. If the child is rationally interpreting evidence of label variation in a *social* setting, we should expect their attention to categorical sound variation to be distributed in accordance with their beliefs about the usefulness of speakers.

I will use the given model of reasoning about categories and speakers of unknown reliability to simulate two experiments from Rost and McMurray (2009), contrasting the behavior of infants habituated to exemplars which were produced by either a single speaker, or by unique speakers. I will then show that a listener who can effectively categorize speech tokens may more confidently reason about the accuracy of an object-label pairing when it is attested by a group of speakers compared to a single informant.

4.4 Modeling Rost & McMurray 2009

In these experiments, infants are habituated to object-label pairings, where the objects are visually distinct and the labels are lexical neighbors (“buk”,“puk”). They are then tested to see if they successfully differentiate same and switch trials - when the labels are applied to either the same object as seen in test (same trial), or to the object which was assigned its lexical neighbor in habituation (switch trial). Our model simplifies the problem by removing several variables - instead of encoding fine phonetic detail, we assume that the infants are capable of categorizing the individual speech tokens. We also eliminate the need to infer the referent of speech tokens, instead supposing that all tokens share the same referent. We will show that even assuming perfect performance on speech categorization and identifying the referent of the speech act, the social inference model predicts the pattern of results seen in Rost and McMurray (2009).

Given some set of observations, the listener must infer both the category and the knowledgeability of each speaker: $P(C, \vec{K_w}|\vec{D_z})$. Increased certainty about the identity of C is expected to correlate with increased surprisal on switch trials and therefore looking time to the target image. By contrast, infants who are unsure of the label should demonstrate low surprisal, with lower looking times.

To model the pair of experiments in Rost and McMurray (2009) we need to do this joint inference under two conditions. In the single speaker condition, all observations are attributed to one source. This inference problem is described in the general case in section 4.4.1. In the multiple speaker condition, each observation is attributed to a unique source. This inference problem is laid out in section 4.4.2. We then return to modeling the experimental data in section 4.4.3.

4.4.1 The single speaker condition

Joint inference of category and knowledgeability

For the condition where the infant hears labels from a single speaker, we model this as a joint inference on (C, K_w) for a sequence of data points $D_{\vec{z}}$. In this condition, all the data points are associated with a single belief about the knowledgeability of the speaker.

$$P(C_x, K_w|D_{\vec{z}}) = \frac{P(C_x, K_w) \sum_y P(D_{\vec{z}}|R_y, C_x, K_w)P(R_y)}{\sum_{x'w'} P(C_{x'}, K_{w'}) \sum_{y'} P(D_{\vec{z}}|R_{y'}, C_{x'}, K_{w'})P(R_{y'})} \tag{4.1}$$

Equation 4.1 gives the joint posterior probability of the category being index x and speaker having knowledgeability w. Assuming the total number of observed data points is m, this probability distribution is defined for three cases, or types of possible observations $D_{\vec{z}}$.

The total number of observed data points must be a sum of the number of ambiguous observations (m_0), and the number of observations unambiguously associated with each possible category x' ($m_{x'}$), as shown in equation (4.2).

$$m = m_0 + \sum_{x'} m_{x'} \tag{4.2}$$

In this model, knowledgeable speakers always have correct intentions (ones which match the true category), so any speaker who produces any data which unambiguously contradicts the listener's belief about the category is deterministically judged to be unknowledgeable. The first section of table 4.1 gives the joint posterior on category and knowledgeability for all sets of observations where $\sum_{x'} m_{x'} > 0$. The total number of categories is given by n, and the expression $T_x(m)$, expanded in equation 4.3 describes the probability of making m ambiguous observations given category C_x and knowledgeability K_w. See Appendix 7.1.2 for a full derivation.

Table 4.1: joint posterior distribution of category and knowledgeability

condition	K_w	$P(C_x, K_w \vert D_{\vec{z}})$
$m > m_0 + m_x$	1	0
	0	1
$m = m_0 + m_x$	1	$\frac{nk}{[1+(n-1)k]}$
$m_0 < m$	0	$\frac{(1-k)}{[1+(n-1)k]}$
$m = m_0$	1	$\frac{c_x k[T_x(m)]}{\sum_{x'} c_{x'} k[T_{x'}(m)] + \frac{(1-k)}{n}[\sum_{x''} T_{x''}(m)]}$
	0	$\frac{\frac{(1-k)}{n}[\sum_{x'}[T_{x'}(m)]]}{\sum_{x'} c_{x'} k[T_{x'}(m)] + \frac{(1-k)}{n}[\sum_{x''} T_{x''}(m)]}$

Both knowledgeable and unknowledgeable speakers select a single intention from which all their subsequent productions are generated. Given the same intention, the specific composition of their datasets is expected to be equivalent, and therefore uninformative with respect to knowledgeability. Therefore, for any set of observations containing at least one unambiguous datapoint consistent with the listener's beliefs about the category the posterior is distributed identically. This is illustrated in the second section of table 4.1; the probability of some category C_x occurring jointly with some K_w is a function of $P(K_w)$ and is identical for all values of x.

In the case where all m observations are ambiguous, the posterior on knowledgeability is distributed over all possible combinations of knowledgeability and category which may produce such a sequence. To simplify this expression, we rewrite the probability of a single speaker producing m consecutive ambiguous data points like so:

$$T_x(m) = \sum_y P(D_0|C_x, K_w)P(R_y) = [(1-r)(1-p_x)^m + r(1-q_x)^m] \tag{4.3}$$

The posterior probability of this speaker being knowledgeable is given in the last section of table 4.1. Only in the case of entirely ambiguous data does the size of the data set influence listener beliefs about speaker knowledgeability. Ambiguous data does not rule out any categories, but increasing amounts of ambiguous data are increasingly good evidence for the category which is most often instantiated in an ambiguous form. Alternately, a single unambiguous data point attributable to a knowledgeable speaker may deterministically rule out all but one category. As a consequence, any amount of unambiguous data from a speaker who has some probability of being knowledgeable will be more informative than any amount of ambiguous data.

If the infants in the experiment are attending to the likelihood that each speaker is knowledgeable, then even supposing they are not encoding the phonetic detail which distinguishes the presented minimal pairs across speakers, we should still expect them to be more sensitive to mismatches between the test object and label when provided with additional evidence that a particular speaker's utterances may be useful. The second experiment in Rost and McMurray (2009), which presented the infants with exemplars from multiple speakers in habituation, provides the infant with just this sort of evidence - detecting agreement in the testimony of multiple speakers could license the inference that each of these speakers is more likely to be knowledgeable than a single speaker's testimony which lacks corroboration.

4.4.2 The multiple speaker condition

Joint inference

For the condition where the infant hears labels from multiple speakers, we model this as a joint inference on $(C, K_{\vec{w}})$ for a sequence of data points $D_{\vec{z}}$. In this condition, each data point is the contribution of a distinct speaker, and as such is associated with a unique belief about the value of that speaker's knowledgeability. In other words, for a set of data with m elements, the listener must now infer a sequence $K_{\vec{w}}$ with length m.

$$P(C_x, K_{\vec{w}}|D_{\vec{z}}) = \frac{P(C_x, K_{\vec{w}}) \prod_i \sum_y P(D_{z_i}|R_y, C_x, K_{\vec{w}}) P(R_y)}{\sum\limits_{x'\vec{w}} P(C_{x'}, K_{\vec{w}'}) \prod_i \sum_y P(D_{z_i}|R_y, C_x, K_{\vec{w}}) P(R_y)} \tag{4.4}$$

The learner must produce one belief about C and a sequence of beliefs about K - the probability of any interpretation $(C_x, K_{\vec{w}})$ is distributed over all possible sequences of data $(D_{\vec{z}})$. This distribution is too complex to calculate analytically, so we will instead use a Markov Chain Monte Carlo method called Gibbs Sampling (Geman and Geman, 1984). Given the conditional distributions, $P(C_x|K_{\vec{w}}, D_{\vec{z}})$ and $P(K_{\vec{w}}|C_x, D_{\vec{z}})$, the Gibbs Sampler iteratively samples from these, using the new value obtained at each step to sample the other conditional distribution. This iterative sampling process will converge to approximate the joint distribution described in equation (4.4). Note that these conditional distributions are modified versions of the Litmus Test and Trust Fall seen in 3.3 and 3.4. See section 7.1 for derivations.

Inferring speaker knowledgeability given category The probability of each speaker's knowledgeability is independent, so the conditional distribution of a group of speakers' knowledgeability values given the category and data is given by the product of the conditional distribution of each individual's knowledgeability.

$$P(K_{\vec{w}}|C_X, D_{\vec{a}}) = \frac{\prod_{w_i} P(K_{w_i}) \sum_Y [P(I_Y|C_X, K_{w_i}) P(D_{Z_i}|I_Y)]}{\sum_{i'} \prod_{w_{i'}} P(K_{w_{i'}}) \sum_Y [P(I_Y|C_X, K_{w_{i'}}) P(D_{Z_i}|I_Y)]} \tag{4.5}$$

The conditional probability of a single speaker being knowledgeable or unknowledgeable is given for each of three conditions in table 4.2 below. The first condition supposes that the observed data D_Z is unambiguously in conflict with the hypothesis that the underlying category is C_X (Z≠X). Because knowledgeable speakers always have correct intentions, any speaker who produces unambiguous data which is inconsistent with the listener's belief about the category is deterministically judged to be unknowledgeable. The second condition supposes that the observed data D_Z is unambiguously supportive of the category hypothesis (Z=X), and the third, that the listener observes only ambiguous data (Z=0).

Table 4.2: posterior probability of K assuming C

condition	w	$P(K_w \mid C_X, D_Z)$
$Z \neq X$	1	0
	0	1
$Z = X$	1	$\frac{k}{[k+\frac{(1-k)}{n}]}$
	0	$\frac{\frac{(1-k)}{n}}{[k+\frac{(1-k)}{n}]}$
$Z = 0$	1	$\frac{k[(1-r)(1-p_X)+r(1-q_X)]}{k[(1-r)(1-p_X)+r(1-q_X)]+\frac{(1-k)}{n}[(1-r)\sum_{Z'} 1-p_{Z'}+r*\sum_Y 1-q_Y]}$
	0	$\frac{\frac{(1-k)}{n}[(1-r)\sum_{Z'} 1-p_{Z'}+r*\sum_Y 1-q_Y]}{k[(1-r)(1-p_X)+r(1-q_X)]+\frac{(1-k)}{n}[(1-r)\sum_{Z'} 1-p_{Z'}+r*\sum_Y 1-q_Y]}$

Inferring category given speaker knowledgeability The conditional probability of the category given a set of m data points, one each from m distinct speakers, each of whose knowledgeability is known, is given by the product of the sum of all conditional distribution of each data point's knowledgeability.

$$P(C_X|K_{\vec{w}}, D_{\vec{Z}}) = \frac{P(C_X)\prod_i \sum_Y P(D_{Z_i}|I_Y)P(I_Y|C_X, K_{w_i})}{\sum_{X'} P(C_{X'})\prod_i \sum_Y P(D_{Z_i}|I_Y)P(I_Y|C_{X'}, K_{w_i})} \tag{4.6}$$

Data from unknowledgeable speakers provides no information about the category, as was illustrated in table 3.9 from section 3.5. Without loss of generality, we may describe the listener as simply ignoring data from speakers who are identified as unknowledgeable. The total number of observed data points from knowledgeable speakers (m_k) must be a sum of the number of ambiguous observations from knowledgeable speakers (m_{k0}), and the number of observations from knowledgeable speakers unambiguously associated with each possible category x' ($m_{kx'}$),, given in equation (4.7).

$$m_k = m_{k0} + \sum_{x'} m_{kx'} \tag{4.7}$$

The posterior probability of the categories given a series of paired values K and D will have three distributions - if no data from knowledgeable speakers is available, then the posterior will be identical to the prior. Supposing any number of knowledgeable speakers give matching unambiguous reports, the listener will deterministically believe these reports. Lastly, if all knowledgeable speakers report ambiguous data, then the posterior on categories is proportional to the likelihood that the observed number of ambiguous reports might occur given each category. See section 7.1.3 for details.

Table 4.3: posterior probability of C assuming K

condition	$P(C_X \| K_{\vec{w}}, D_{\vec{z}})$
$m_k = 0$	c_x
$m_k = m_{k0} + m_{kx}$	1
$m_k = m_{k0}$	$\frac{[T_x(1)]^{m_{k0}}}{\sum_{x'}[T_{x'}(1)]^{m_{k0}}}$

4.4.3 Simulation

The participants in the Rost and McMurray (2009) Experiment 1 heard seven consecutive instances of the same exemplar. We simulate the beliefs of the infant at the end of this habituation period by using our model to calculating the joint posterior probability of the category and the speaker's knowledgeability after seven instances of matched unambiguous tokens. This simulation supposes a three-way phonetic contrast, rather than the two-way contrast in the original experiment.

This adjustment is reasonable, considering that it is not possible to know how many potential categories infants may initially hypothesize, and the more complex problem biases the simulation against our hypothesis by making knowledgeability more difficult to infer. The contrast used in the experiment ("buk," "puk") represents the only phonemic contrast defined by VOT in English phonemic, but three-way and even four-way VOT distinctions are attested in many languages (Lisker and Abramson, 1963;Lisker and Abramson, 1964). Given the absence of ambiguous data, the listener's posterior distribution over knowledgeability will be the same for all categories, and a function of the prior on knowledgeability and the number of categories. The model's predictions will pattern the same no matter how many categories there are - the more corroboration that a speaker's utterances have, the more likely they are to be knowledgeable.

Instead of seven exemplars belonging to the same voice, the subjects in Experiment 2 heard the same number of exemplars, each produced by a different speaker. The authors argued this additional variation in phonetic data was responsible for the difference in results between Experiments 1 and 2. However, the model used here assumes that the infant's phonetic categorization is fully functional, and does not distinguish between different levels of phonetic encoding.

In both experiments, the infants' sensitivity to the subsequent presentation of a contrasting stimuli was measured in the amount of time the infants spent looking at the visual display. As an analog, we will compare the model's predictions for what the child believes about C and K in the two conditions. The more surprisal associated with an event, the higher the predicted looking time.

Experiment 1

The joint posterior probability on category and knowledgeability was calculated using the analytically derived joint probability distribution from section 4.4.1. The probability that the infant believes speaker is unknowledgeable after seven presentations from one speaker in a simulation with three possible categories is calculated to be 25%. This figure is the sum of the last three lines in table 4.4.

We estimate the relative entropy associated with a same and a switch trial by calculating the probability of the the event where C has an identity that does or does not match the exemplars given in the habituation phase, and find that these sums are equivalent. The model therefore predicts that infants will not attend longer to the switch trials.

Table 4.4: subject beliefs post-habituation: Experiment 1, word 1

condition	$P(C_x, K_w \| D_{(1,1,1,1,1,1,1)})$
$C_x = 1, K_w = 1$	$\frac{1}{2}$
$C_x = 2, K_w = 1$	0
$C_x = 3, K_w = 1$	0
$C_x = 1, K_w = 0$	$\frac{1}{6}$
$C_x = 2, K_w = 0$	$\frac{1}{6}$
$C_x = 3, K_w = 0$	$\frac{1}{6}$

Experiment 2

The joint posterior probability of category and knowledgeability was estimated using the Gibbs sampler described in section 4.4.2. The results of a chain with 1000 cycles are given in table 4.5. The probability that the infant believes the next utterance will be from an unknowledgeable speaker after seven presentations from seven different speakers remains at 50%, unchanged from the first experiment. We use the posterior probability of the familiarized label being correct to predict surprise on the switch trial. The equation for this posterior is given in equation 4.8.

$$P(C, K_1, K_2, K_3, ..., K_m | D_1 = D_2 = ... = D_m = buk) \propto \prod_i P(D_i | C, K_i) P(C, K_i) \tag{4.8}$$

Table 4.5: subject beliefs post-habituation: Experiment 2, word 1

condition	$P(C_x, K_w \| D_{(1,1,1,1,1,1,1)})$
$C_x = 1, K_w = 1$	0.9988
$C_x = 2, K_w = 1$	0
$C_x = 3, K_w = 1$	0
$C_x = 1, K_w = 0$	0.0004
$C_x = 2, K_w = 0$	0.0004
$C_x = 3, K_w = 0$	0.0004

The model predicts that the listener will interpret the speaker's unambiguous agreement with the other speakers to indicate the speaker is overwhelmingly likely to be knowledgeable. Accordingly, the switch trials, are expected to be more surprising to the infants, and receive more looking than the same trials.

4.4.4 Discussion

Previous interpretations of this task have attributed 14-month olds' failure to lexical processes that prevent the infant from attending to fine phonetic detail, to the absence of necessary referential cues, or to a lack of sufficient meaningful variation in the source. Instead, it is possible that the observed effect arises from attentional preferences for social partners. Infants who fail on the Switch task may be demonstrating more selective preferences for linguistic informants.

The failure of an infant to recover attention to the new stimulus on a switch trials implies a failure to discriminate between the two stimuli. However, the existing literature does not rule out a hypothesis where listeners performance at phonetic discrimination is correlated with beliefs about informant reliability.

Although Rost & McMurray interprets their pattern of results as evidence of infants relying on additional “bottom-up” phonetic variation to learn words, the foregoing simulations demonstrate that the improvement shown by 14-month olds in the multiple speaker condition can also be explained through the use of “top-down” heuristics relying on judgments about the nature of the source. Under the assumption that knowledgeable and unknowledgeable speakers are uniformly distributed, data corroborated by multiple speakers provides a clear potential advantage to the listener. The more likely a randomly selected speaker is to be reliable, the less additional information is expected from sampling additional informants, predicting that perception of agreement among informants will increase the infant’s surprisal at conflicting information. Improvement on the Switch task may therefore coincide with the infant developing more sophisticated strategies for determining the reliability of linguistic informants.

Associating an object with a label also requires the coordination of other cognitive processes, including attention, segmentation and inference about the speaker's referential intent. The failure to demonstrate phonemic discrimination on the Switch task has sometimes been attributed to a resource limitation (Stager and Werker, 1997; Pater, Stager, and Werker, 2004). However, these studies attempted to measure infant recognition of words assuming that the representational structures at issue did not reflect any meaningful variation in encoding of details about informants or their reliability. We will now consider results from other investigations using the Switch task, and examine the implications of the social inference model for the interpretations of their findings.

Familiarity effects In tasks involving familiar words and objects, 14-month-olds demonstrate increased sensitivity to phonetic detail (Swingley and Aslin, 2002; Fennell and Werker, 2003; Fennell and Werker, 2004; Fennell, 2012). Supposing the infants' behavior is attributable to the increased task requirements of the audio-visual associative learning required to respond to novel words, then the presentation of familiar stimuli should alleviate that difficulty. In effect, the participants' a priori word knowledge appears to facilitate the task.

In our model, we can simulate this contrast by increasing the prior on knowledgeability. The parameter K in our model predicts the likelihood of an informant both correctly identifying and labeling the referent, which, whether familiar or novel, is known to the subject. Assuming that the child believes that a familiar object is more likely to be known to their interlocutor, or that this object is simply more salient, an increase in P(K) simulates the effect of familiar stimuli. Rather than the familiarity of the lexical items facilitating lexical processes, it may simply facilitate epistemic trust in the informant, resulting in greater phonetic sensitivity.

Referential ambiguity Other studies demonstrate that the infants who fail on this task see improved performance when additional referential cues are present. Performance on the Switch task improves when the novel word is embedded in an overtly referential phrase (i.e. ”look at the *blick*”) (Fennell and Waxman, 2006), when the training phase contains familiar named objects (Fennell, Waxman, and Weisleder, 2007). However, when familiar objects in habituation are paired with exclamations (e.g., ”Wow!” or ”Whee!”), no improvement is observed (Fennell and Waxman, 2010). These results support the hypothesis that 14-month-old infants’ failure on the Switch task is a consequence of referential ambiguity. Cues which make the stimulus presentation more clearly a referential act increase the likelihood that infants demonstrably create a mapping between the word and object using fine phonetic detail.

However, the task is already designed to make the labeled object salient to the subject. Assuming, as we have before, that the subjects do know which object is being referred to, the inclusion of additional cues that the speech act references this object may again be encoded as an increase in the prior probability on K, ascribing a greater likelihood to informant knowledgeability. Rather than simply tracking the speech acts themselves, a listener who is also sensitive to source may interpret additional referential cues as a reflection on the quality of the linguistic informant. We expect any stimuli which biases the infant to believe the informant is more likely to select both the correct referent and label will also result in an increase in surprisal for switch trials, and consequently improved performance on the task.

14-month olds perform above chance on a preferential looking paradigm, suggesting that the difficulty observed under other conditions could be the result of task difficulties (Yoshida et al., 2009). Supposing that the subject assigns the event that the speaker heard in the habituation phase is knowledgeable a non-zero probability, then the model predicts a preference for the labeled object during a preferential looking test. This slight preference is also predicted by the model given in Apfelbaum and McMurray (2011).

Investigating preferences for informants with the switch task In order to explore the source tracking hypothesis, which links processes of epistemic trust and performance on phonetic discrimination tasks, several modifications to the task may be useful.

Firstly, the source-tracking hypothesis predicts that 14-month olds would demonstrate improved phonetic sensitivity on a Switch task featuring novel words recorded by familiar speakers. However, such a result would also be compatible with a cognitive load hypothesis.

The task might also be modified to contrast presentation of novel words in carrier phrases recorded by speakers in alternate dialects or languages. If foreign-language carrier phrases provide any benefit then this would challenge the source tracking explanation. However, infant sensitivity to within-language dialects would not necessarily be inconsistent with this hypothesis.

Pre-test trials showing the speaker labeling familiar objects improve the performance of 14-month olds (Fennell and Waxman, 2010). Fennell and Waxman interpreted these results to indicate that the infants were responding to presence of additional cues that the speech was referential in nature. However, this study was done using stimuli recorded in a single voice. It is possible that the pre-test trials were interpreted by the subjects as evidence that the speaker is a credible source of linguistic data, and the improvement in performance on trials where the child witnessed accurate labeling behavior is actually an effect of epistemic trust formed during the habituation.

Suppose a habituation featuring labeling from two speakers. The source-tracking hypothesis predicts that whether children demonstrate sensitivity to a phonetic contrast will be predicted by their belief that the speaker is knowledgeable. Supposing one of the speakers heard in pre-test is more reliable at labeling familiar objects, infants who hear this speaker's voice on test trials should be more likely to attend to switch trials than infants who hear the less reliable speaker's voice at test.

Likewise, the use of a pre-test demonstration where the speaker is shown to be more or less reliable using non-linguistic cues (such as indicating with gaze where an object will appear) may diminish the beneficial effect of naming familiar objects pre-test. If infants are attending to the reliability of the speaker, then demonstrations that they are unknowledgeable in other ways may cause the infant to disprefer attending to that informant's phonetic variation.

Implications for the growth of the lexicon There is some evidence that performance on the Switch task is correlated with vocabulary size (J. F. Werker et al., 2002). The source-tracking hypothesis predicts that children who perform poorly on this task may be demonstrating higher informant selectivity. A child who selectively attends to a smaller number of speakers will have a smaller less diverse data set - they will process fewer linguistic tokens and less variety of tokens than a child who attends to a larger number of speakers. This suggests that lower vocabulary sizes may be a direct effect of children having formed narrower preferences for linguistic informants.

Summary The model presented here provides a united explanation for infants' pattern of performance on multiple variations of the Switch task. While the traditional model of language acquisition predicts that infants will learn more from sources to whom they are more frequently exposed, the source-tracking hypothesis predicts that infants will prefer the label offered by the majority of pre-test speakers, excepting when they have observed evidence that those speakers are unreliable.

To provide confirmation of this hypothesis, it is necessary to conduct a systematic comparison of infant performance after exposure to different amounts of testimony from differing numbers of informants. It is also necessary to determine how allocation of epistemic trust may vary between populations. Children from different cultural backgrounds and learning in different modalities are expected to eventually acquire distinct strategies for determining the reliability of an informant. Therefore, before we may tease apart the effects of exposure and epistemic trust on word learning, we must understand normal variation in its application. The present work suggests a new research program uniting studies of developmental social psychology with psycholinguistic processing, to discover how variation in phonetic representations are affected by the perception of identity, including attributes such as authority, gender and race.

In the next chapter I will review evidence of the links between adult phonetic and social perception, and then use the present model to simulate a study of infant behavior in response to labels from different types of linguistic informants. The results suggest that preverbal infants are sensitive to perception of linguistic informants' social group membership, implicating a role for perception of identity in early word learning.

Chapter 5

Word Learning with Beliefs about Speakers

The goal of this chapter is to investigate what effects the processes that infants use to categorize informants may have on the phonetic perception of pre-verbal infants and their acquisition of early lexical items. Put another way, can perceiving information about *who* is speaking impact *what* infants believe is said? In the previous chapter I presented possible evidence that 14-month-old infants use measures of epistemic reliability to guide their selection of language informants. In other words - infants may categorize informants as either knowledgeable or unknowledgeable, and divide their attention accordingly. Do infants use judgments about the membership of informants in categories and epistemic judgments about those categories to make predictions about the reliability of speech?

In this chapter, I will explore how perception of other speaker cues may indirectly affect infant expectations of language use by influencing the perception of informant knowledgeability, showing evidence that preverbal infants do preferentially attend to kinds of speakers which they expect to be more knowledgeable.

I will begin by describing variation in sociolinguistic competence among adults to aid in broadly defining the trajectory of sociophonetic development, before returning to a discussion of infant social perception. Next, I will describe an example of how beliefs about group membership can be leveraged to identify reliable sources, adapting our model to simulate an experiment from Koenig and Echols (2003). Looking time patterns suggest that infants are relying on knowledge about the group membership of the speech informant to make inferences about both the quality of the speaker and their speech.

I argue that the literature supports the hypothesis that infants make inferences about membership in socially defined groups which in turn impact their expectations of informant knowledgeability. Both of these inferences are therefore expected to have an a priori impact on beliefs about the reliability of both linguistic and non-linguistic informant behavior. Rather than identifying correct speech patterns independent of listener, language users must be able to execute two related tasks - matching speech patterns to informants, and evaluating those patterns for correctness. I will show that differing expectations about the reliability of groups necessarily predict that under uncertainty about knowledgeability, listeners must consider competing linguistic standards when evaluating labels.

Lastly, I will review some evidence that infants may rely on physical features to make these inferences. I will argue that in addition to patterns of speech perception and production which are broadly characteristic of their language, learners also tend to acquire a pattern of *speaker perception* which is characteristic of their social group. I use "speaker perception" here to reference two kinds of perception - first, linguistic speaker perception, or variation in the speech signal which is indexical (indicating the identity of the speaker). This variation in speech is neither arbitrary nor idiosyncratic, but is systematically related to the perception of affiliation with social groups, such as those defined by geographic origin or socioeconomic status. This knowledge forms the sociophonetic basis for the perception and performance of social identities.

Second, learners also develop non-linguistic speaker perception, acquiring beliefs about non-linguistic features and behaviors which distinguish people with distinct social identities. Tracing the development of adult sociophonetic knowledge will require uniting accounts of how children attend to both linguistic and non-linguistic cues as potential markers of affiliation.

5.1 Adult speech and speaker perception

Adult-like phonetic processing is characterized by categorical perception, with stronger discrimination of between-category contrasts than within-category contrasts (A. M. Liberman, Harris, Hoffman, and Griffith, 1957). Importantly, it is between-category contrasts which indicate differences in the referential content of utterances made by native speakers. For example, the phonological difference between "pin" and "bin" distinguishes these words as referring to different kinds of objects. However there are also systematic differences between how such categories are deployed in different speech varieties. For example, Spanish accented speakers of English are more likely to pronounce the aspirated $[p^h]$ as a non-aspirated $[p]$. Evidence shows that adult listeners display a remarkable sensitivity to dialectal variations in phonetic distributions. This natural sensitivity has virtually ubiquitous implications for the outcomes of social interactions with other language users.

For example Purnell, Idsardi, and Baugh (1999) describes the result of an apartment search conducted over the telephone in different neighborhoods of San Francisco, CA. All phone calls were made by a single investigator in one of three ethnically coded dialects. It was found that the likelihood of the investigator making a successful appointment to view an apartment depended on whether the dialect he used matched the predominant ethnicity of the neighborhood where the apartment was located. The results of the behavioral study were supported by perceptual identification experiments using the single word 'hello.' Participants showed statistically significant abilities to identify ethnically affiliated dialects from very short samples of speech. Subsequent fMRI studies suggest that these dialect categories are accessed early in automatic pre-attentive speech processing (Schachner and Hannon, 2011; Tuninetti, Chládková, Peter, Schiller, and Escudero, 2017).

William Labov found that the usage of the phoneme /r/ differentiates social class in New York city speech (Labov, 2006). He found that staff at higher-class stores were more likely to pronounce the /r/ sounds in the phrase "fourth floor," especially in an emphatic repetition. However employees at stores more often frequented by less affluent customers were more likely to produce r-less speech. These differences in production correlate with perceptible social categories, and therefore function as affiliative cues. I will use the phrase *linguistic speaker perception* to refer to inferences about the membership of the speaker according to observation of their speech.

Traditional models assess the value of linguistic behavior with respect to a single standard, describing a single community of practice. However, the perceived quality of referential content in a labeling action may depend both upon the word form provided and upon the affiliation of the linguistic informant who produced it. In this model, speakers who belong to a group which is expected to be comprised of useful linguistic models are expected to produce accurate and not inaccurate word forms, while speakers who belong to a group which is less knowledgeable are expected to potentially produce both accurate and inaccurate word forms.

A speaker who is suspected to be knowledgeable is therefore expected to produce speech will be valuable for two reasons: it will predictably both contain accurate word forms and meanings, and it will fail to contain inaccurate, or deviant word forms. Conversely, the unknowledgeable speaker is expected to produce speech which is unhelpful for one reason: it is unpredictable. In other words, although the word forms produced by an unknowledgeable speaker may sometimes match those of knowledgeable speakers, they cannot be expected to do so in a consistent way. However, knowledgeable speakers are expected to exclusively produce intentions which correspond to those of other knowledgeable speakers.

5.1.1 Modeling labeling with speaker perception

Hoowever, a listener's perception of a labeling instance is not only dependent upon their beliefs about how the label is generally produced, but also upon beliefs about how different speakers will produce that label. In addition to the perception that for an object, some labeling behaviors are more valid than others (e.g. in labeling Fido a speaker may judge "dog" /dɑːg/ is more acceptable than "dawg" /dɔːg/, while both "dot" /dɑːt/ and "dawt" /dɔːt/ are unacceptable), listeners also acquire the belief that some labelers habitually perform more standard labeling behaviors than others, consequently perceiving some speakers as accented and others as unaccented. Further, listeners are capable of forming linguistic expectations based on non-linguistic attributes of a speaker, including their regional origin (Niedzielski 1999) age and socioeconomic status (Hay, Nolan, and Drager, 2006).

While adults have robust and consistent judgements about the patterning of both within and between category phonetic contrasts and the affiliative nature of these patterns, these beliefs are not generally considered to be part of lexical knowledge. It is critical to understand how perception of speaker affiliation develops in conjunction with a phonology which supports the interpretation of referential content. What is the relationship between speech perception and speaker perception and how are the developmental trajectories of these two systems distinct? I will show that there is evidence that infants have expectations about the linguistic behavior of speakers based on perceptions of the speaker drawn from both linguistic and non-linguistic data.

I use the phrase ”non-linguistic speaker perception” to encompass the processing of any non-linguistic signals which may indirectly inform the listener’s expectations regarding referential linguistic behavior by impacting their expectation of the speaker’s linguistic behavior. In this chapter we will consider two kinds of non-linguistic speaker perception - that based on information access, and that based on affiliation. While lexical items may ostensibly be interpreted independent of speaker perception, I will show evidence that even in infants, the listener’s evaluation of a lexical item displays sensitivity to differences in both distributional phonetic patterns among knowledgeable informants and the distribution of knowledgeable informants themselves.

In order to learn the meaningful variation in speech language users must preferentially attend to kinds of speakers and their *characteristically correct* variation, which requires differentiating these speakers both from each other and from those whose variation is unpredictable. In this chapter I will show evidence that preverbal infants recruit both the non-linguistic cues of information access and evidence of affiliation with similar speakers to make categorical judgements about the identity and validity of labels provided by a given source.

5.1.2 Speaker perception and affiliation

Adult interpretation of spoken language provides cues to many socially valuated characteristics, such as an individual's ethnicity, regional background and social class (Labov, 1991; Labov, 2006). Adults exposed to audio-only samples of content-neutral speech are able to use sociophonetic cues to interpret a speaker's identity (Remez, Fellowes, and Rubin, 1997), regional dialect (Clopper and Pisoni, 2004), and ethnicity (Purnell et al., 1999) with accuracy greater than chance. We may *also* interpret these evaluations of identity based on speech to indicate that listeners perceive speakers to be affiliated with specific abstractly defined dialectal communities.

In other words, we might interpret a listener's reported perception that a speaker is female and African-American as a judgement that this speech indicates membership in a female speech community, an African-American speech community and an African-American female speech community, with each of these judgements relying on some combination of cues extracted from the speech signal. The stable perception of social categories from exposure to voice samples demonstrates the ability to generalize that features found in these voices may also be found in other members of those social categories. Furthermore, speakers of a target dialect not only make consistent judgements about what speech might indicate given identities, they also make consistent judgements about the relative reliability of speakers with these identities (Frumkin, 2007). The challenge of language acquisition is thus not limited to recovering the structure of the native language, but also conventional beliefs about sources, their reliability and their membership in social groups.

We will therefore refer to any information associated with an informant which provides the basis for learners to make a socially based categorization as an *affiliative cue*. Affiliative cues license the perception that a speaker is associated with other informants belonging to an abstract group. Group membership is simply a way of encoding the prediction that speakers who share a characteristic will exhibit similar behavior, so that we may refer to groups defined by linguistic behavioral expectations, non-linguistic behavioral expectations or both. Supposing that groups have different characteristic levels of knowledgeability, we should expect learners' developing beliefs about groups to be apparent in measures of attentional preference.

If infants are aware that some groups are characteristically more knowledgeable than others, we should expect perception of features which reliably indicate an informant's affiliation with a more knowledgeable group to induce greater attention to labeling behavior than perception of features which suggest affiliation with a less knowledgeable group. In other words, speech informants who are a priori expected to be more accurate labeling sources should receive more attention from infants than speech informants who are expected to be inaccurate. Importantly, I an predicting that this expectation holds for both the speaker themselves and their utterances - a listener's attention to a speaker will be greater for a speaker who is expected to be reliable, consequently impacting the listener's attention to both linguistic and non-linguistic aspects of the stimuli they provide.

Listeners are also sensitive to non-affiliative cues which indicate that a particular speaker can access relevant information, or has *perceptual access*. For example, 8-month-old infants preferentially follow the gaze of an informant when they have previously observed that this informant's gaze is a reliable cue to the appearance of an object (Chow, Poulin-Dubois, and Lewis, 2008). To the extent that affiliative cues indirectly predict evidence of knowledgeability, they may also potentially impact listener responses to perceptual access cues. In the next section, I will present evidence that infant speech perception is constrained by both non-linguistic and linguistic expectations induced by listeners' speaker perception.

5.2 Child and infant perception of speech and speakers

Early in the process of constructing representations of what their language is, infants demonstrate a sensitivity to the nature of a speech source in their phonetic learning. In this section I will review evidence that infants are attentive to non-linguistic attributes of linguistic sources, and explore two types of cues to knowledgeability - affiliative cues, which indirectly indicate general knowledgeability, and perceptual access cues, which directly implicate a speaker as knowledgeable in a specific context. I attempt to answer the question: what generalized beliefs about informants, reliability and membership in social groups do children come to hold, and how might these beliefs affect the development of the lexicon?

5.2.1 Non-affiiative cues to information access

Chow et al. (2008) exposed 14-month-old infants to an informant who expressed happiness as they looked inside a container that either held a toy (reliable looker) or was empty (unreliable looker). Infants were then given the opportunity to follow the same informant's gaze to a target object located either in front of or behind a barrier. In the test trials, infants followed the gaze of speakers in both conditions to target objects that appeared in front of a barrier, suggesting that they treat the cue of speaker gaze as a source of referential information. However, only infants in the reliable condition consistently followed the gaze of unreliable lookers to objects behind the barrier. These results demonstrate that at 14 months, infants preference to follow a speaker's gaze is a product of both experience with a specific speaker and the generalized prior belief that speaker gaze contains referential information. These results suggest that infants are more sensitive to the intentionality of behaviors performed by speakers who they have observed to be epistemically trustworthy.

There is other evidence that at this age, infants are forming beliefs about the intentionality of actions. In Carpenter, Akhtar, and Tomasello (1998), 14-18 month old infants watched an adult interact with an object, producing an interesting result. When given the opportunity to reproduce the outcome, infants were half as likely to imitate actions which had been accompanied by an exclamation indicating an accidental outcome (”oops!”) compared to those which were marked as intentional (”there!”) (Carpenter et al., 1998). By this age there is also evidence that infants will use adults' pointing gestures, not just to guide their looking behavior but to infer that the adult is directing them to find a hidden toy (Behne, Carpenter, and Tomasello, 2005; Behne, Liszkowski, Carpenter, and Tomasello, 2012)

5.2.2 Non-linguistic affiliative cues

In the last chapter, I showed evidence that 14-month-old infants attend to source quality, but there is evidence that even younger infants are recruiting source tracking to guide their phonetic learning. For example, P. K. Kuhl (2007) exposed 6-8 month old English learning infants to Mandarin, either through a televised source or from a live social partner. Those infants who received exposure through the electronic display did not show learning on a non-native contrast, while those who were exposed to a live social partner showed robust and durable perceptual learning. The results suggest that socially based perception has a significant impact on phonological learning before the second year of life.

In an experiment with older infants, Spokes and Spelke (2017) showed that 15-18 month-old infants expected two adults who comforted the same baby, or two babies comforted by the same adult to show affiliative behavior, suggesting that at this age infants are making generalizations about social groups based on caregiving relationships. Jin and Baillargeon (2017) compared how infants responded to displays of models who had either previously affiliated or not, subsequently engaging in helping behavior or failing to do so. Infants looked longer to the displays of previously affiliated models ignoring in-group members who needed assistance, demonstrating that at 17-18 months-old infants expect in-group members to help one another (Jin and Baillargeon, 2017). Both of these experiments show that infants perceive models' membership in groups, and make generalizations about their behavior based on that perception.

There is some evidence that infants are forming beliefs about social groups earlier in development. For example, L. J. Powell and Spelke (2013) showed that 12-month old infants treat certain non-linguistic behaviors as affiliative, expecting characters who are shown socializing together in a pre-test familiarization to exhibit similar behavior to one another during test. What is the impact of these beliefs on linguistic perception?

I will argue that infant behavior in response to labels reflects beliefs about both the affiliation and the intentions of informants as speech agents. Specifically, the belief that a label will be accurate is predicated on the belief that the source is likely to be epistemically trustworthy. To the extent that the belief that a label is accurate is based on the belief that the informant belongs to a latent social category, or group, we should expect them to have an a priori expectation that the source's behaviors will reflect some proportion of epistemically trustworthy intentions.

In the next section I will review evidence that the linguistic behavior of preverbal infants reflects inferences about social categories. I will then model an experiment from Koenig and Echols (2003), demonstrating that prior expectations about the knowledgeability of sources according to their membership in groups predict the attentional pattern of 16-month-old infants' in their responses to true and false labeling events.

5.3 Social categories and infant perception of labeling

By 9 months, infants already have the expectation that informants who speak the same language belong to the same social group (Z. Liberman, Kinzler, and Woodward, 2014). It has been shown that the presentation of labels with objects aids basic-level categorization for 9-month-olds (Balaban and Waxman, 1997) and global-level categorization for 12 to 13-month-olds (Waxman and Markow, 1995). However infants do not show evidence of categorization when objects are presented with non-speech or non-labeling sounds (Xu, 2002; Fulkerson and Haaf, 2003). These findings suggest that preverbal infants are strongly predisposed to attend to referential information in speech. However, do infants have the same expectations about referential information for all speakers? How do different beliefs about the reliability of informants impact infant beliefs about the meaning of labeling actions?

Fulkerson and Haaf (2003) also showed that 15-month old infants were able to achieve global categorization of objects when the labels they heard were produced orally by an experimenter, but not when the labels were produced by a tape recorder, suggesting that between 9 and 15 months infants are already using inferences about the value of different sources to guide their attention to labels. In the next section I will model a similar study, Koenig and Echols (2003), which measured infants' looking time to labeling sources and objects for both a human and non human labeling source. The results are consistent with the hypothesis that infants believe that an informant's speech has more value as a consequence of their membership in the category of human informants. In the next section I will review a study of accent perception and discuss how it begins to answer the question - what impact, if any, do beliefs about affiliation with social groups have on early word learning?

5.3.1 Infant perception of accented labels

Children acquiring language are routinely exposed to varieties of speech which do not match their home dialect. For example, a 2015 US Census Bureau report estimates that over a quarter of the US population speaks a language other than English at home, and of those, 40% of those respondents describe themselves as less than proficient in English (United States Census Bureau, 2015). This figure represents an underestimation of the variation available to children, as it does not describe within-language variation in use of linguistic items, registers, styles or vernacular dialects. Whether or not a child is learning to produce multiple dialects, as part of the acquisition process children must implicitly distinguish between multiple speech varieties. How do children learn to cope with non-native speech varieties?

Mulak, Best, Tyler, Kitamura, and Irwin (2013) contrasted the looking preferences of 15-month-old Australian English learning infants when hearing labels produced in either their native accent, or in an unfamiliar Jamaican English accent. The authors argue that mature interpretation of both the native and accented forms of the label should reflect the belief that both forms indicate the same referents, independent of phonetic variation. However, this defines no role of speaker perception in the interpretation of the labeling action. A more accurate discription of adult behavior on this task would need to account for the formation of social judgements about the speaker in addition to an interpretation of the referential content.

If infants are aware that the unaccented token is likely to be a vastly superior representation owing to its function as an affiliative cue, then we should expect them to attend to the unaccented token in more detail and show more evidence of recognizing the label in that condition. In other words, if infants have the belief that native speakers are characteristically reliable, then we should expect them to sustain attention to these speakers more readily, independent of any direct observation of their speech.

Likewise, if the unfamiliar, accented token fails to convey expected information about group affiliation, then perception of the speaker as unknowledgeable may inhibit attention to the labeling action. Cues which indicate membership in accented speech communities are therefore predicted to preemptively inhibit attention to both indexical and referential phonetic variation. Knowing that a speaker is likely to be reliable should cause infants to expect that speaker to produce not just reliable but iconic speech.

Earlier in development infants appear to already have beliefs about the amount of attention a source requires according to their speech accent. Infants appear to preferentially imitate models in non-linguistic tasks when they have seen evidence that the model is a knowledgeable member of their social group (Poulin Dubois et al., 2011; Zmyj et al., 2010). This preference for epistemically reliable models, particularly from the linguistically defined "in-group" is evident in the first year of life, well before infants begin to produce words and suggests a belief that not just epistemically trustworthy actions, but *intentions* are characteristic of that group. I predict that infants will interpret labels in accordance with rational expectations both about quality of the informant and the expected quality of their speech. Beliefs about the prevalence of knowledgeability in different populations will lead to both an increased a priori trust in informants who are presumed to be members of a more knowledgeable group, and decreased trust for those presumed to be affiliated with a less knowledgeable one.

Indeed, 15-month olds only showed a significant preference for the matching picture when the label they heard was produced in their native accent, and not when the label was produced in an unfamiliar Jamaican accent (Mulak et al., 2013). The authors interpreted these results with the assumption that the two distinct forms were equally salient and could be perceived in equal detail. It was therefore suggested that the bias infants showed was a phonological overspecification, which causes the unfamiliar form to be interpreted as a distinct label. Therefore, by 19 months, when infants reliably attend to the same referent for both the accented and unaccented labels, they have ostensibly achieved *phonological constancy*, recognizing that systematic phonetic variation in a word can violate native-accent phonemic boundaries without changing the identity of the word. However, constant with respect to what?

The more adult-like pattern shown by 19-month-old infants reflects a belief not just that the labels produced belong to the same referents, but that the Jamaican and Australian accented English speakers belong to the same speech community. The assumption is that this adult-like perception of both accented speakers having mutual membership in the infant's speech community is veridical. Ignoring accent, both speakers are in fact, speaking English. However, insofar as the 15-month-old infants treat these two labels as distinct, this may simply be reflecting an equally accurate but more precise belief about the talkers' membership in accented speech communities, only one of which is shared with the infant. Only the unaccented speaker speaks the infant's target dialect.

The assumption that infants need to acquire only a single common phonological representation which is dedicated to interpreting referential and not indexical variation is perhaps logical in a study of mono-dialectal children. However, infants who are exposed to significant dialectal variation within their own speech community have distinct challenges. The presence of indexical cues which indicate an affiliation with an unfamiliar speech group may suppress word learning by decreasing attention to detail, and increasing the perception of ambiguity in the referential content. Affiliation with an unknown group will necessarily inspire less selective trust than affiliation with an in-group. Affiliative cues, coupled with preferences for same therefore provide a framework for interpreting speech relative to a standard which is dominated but not exclusively defined by that in-group.

Infants who prefer the matching picture when hearing an accented label are successfully inferring a single meaning despite not just unfamiliar phonetic variation, but specifically phonetic variation which cues belief about membership in distinct social groups and therefore relative unreliability. Rather than developing a single language-wide phonology for interpreting strictly referential content, infants may instead be developing the cognitive control to preferentially attend to the referential content of speech despite the interference from non-native affiliative cues.

Failure to acquire the correct meaning for the accented label is possibly evidence of a learning strategy which supports the acquisition of the native accent. Rather than assume infant behavior reflects a judgement about only the referential content we hypothesize that the behavior is consistent with a more nuanced judgement which reflects beliefs about the certainty with which this judgement can be made. In effect, native speech forms provide a more reliable basis by which to infer the intentions of speakers.

By contrast, accented speech is predicted to induce a belief that the speaker is likely unreliable, which is predicted to reduce attention to referential cues in their speech independent of the speech form provided. Attending to accented speech is therefore predicted to require more time and cognitive control than processing unaccented speech, and functional definitions of accentedness will necessarily depend on the specific language experience of the child. Additionally, as language experience grows, increased processing speed and accuracy at identifying word forms is predicted to grow as a function of exposure *to the target dialect* as spoken by *trusted speakers*.

This account suggests that in the second year of life, infants are becoming less sensitive to acoustic markers of in-group membership, and more willing to interpret non-standard phonetic variation in labels as such. The infant's sensitivity to the sociophonetic accuracy of labeling behavior coupled with their robust beliefs about affiliative behavior suggests that preverbal infants are aware of linguistic behavior both as an affiliative expression, and as an expression of epistemic knowledgeability.

5.4 Modeling infant expectation of labeling accuracy

Koenig and Echols (2003) compared the responses of 16-month old infants to true and false labeling events provided by different kinds of sources. They found that in response to incorrect labeling events, these infants showed longer looking times to the source of the label when it was a human experimenter. However, they did not look longer to an audio speaker when it was the source of the incorrect object labels.

In this scenario, the infants attend to the type of informant: whether it is an experimenter or an inanimate speaker, and accordingly have differing expectations about whether the informant will be a reliable source of object labels.

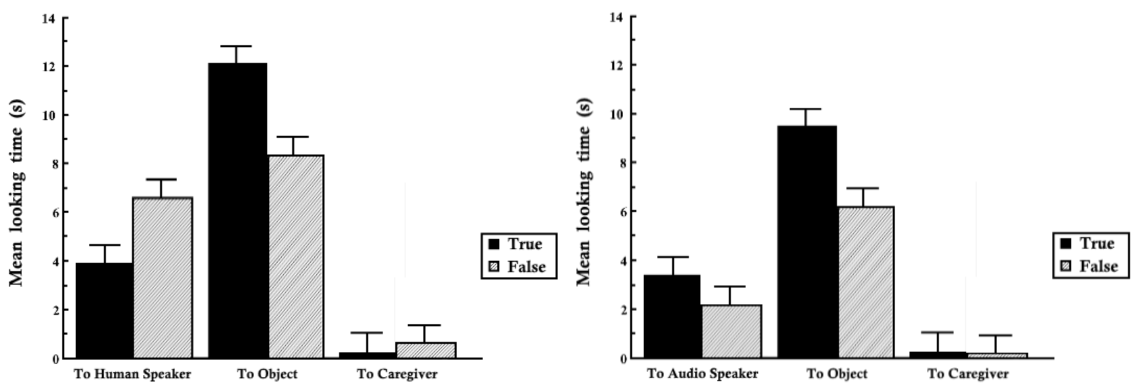
In Koenig and Echols, 2003, the experimenters presented 16-month old infants photographic color slides of five familiar objects: a chair, duck, cat, ball, and shoe. As each image was displayed, the informant provided a label for it by reporting "That's a ____" In the control condition, all of the labeling events were correct, matching the displayed objects. In the test condition, the all of the labels were false. For example, while a picture of a cat is displayed the infant might hear "That's a shoe." Researchers coded the infants behavior, measuring the amount of time they spent looking to both the displayed object, the sources of the labels, and to their caregiver.

In experiment one, the infants heard the labels from a human experimenter seated next to them. In experiment two, the labels were provided by an audio speaker placed in the same location. The researchers hypothesized that the infants' attention to the source would be influenced both by the accuracy of the label and the type of source providing it.

Indeed, they found a broad effect of label accuracy: the infants looked longer at the object when hearing true labels, than when hearing false labels. They also found an effect of label source: the infants looked longer to both objects and label sources when the label sources were human speakers than when they were audio speakers. Lastly, there was an interaction of these two effects: infants looked longer to their caregivers and to the human speakers when labels were false rather than when they were true, but within the audio speaker condition, their looking to the label source was not significantly affected by accuracy.

Overall, infants looked to more to both the object and the speaker in the human labeler condition, as shown by the total area of the bars in the graph on the left of figure 5.1. Comparatively, the total looking time to both the labeler and the object is lower in the audio speaker labeler condition. Provided some certainty that the labeling source belongs to a group which is likely to be unknowledgeable (audio speakers), I will show that a Bayesian model predicts the infant should find correct labeling events from this source more surprising than incorrect labeling events from this source. Likewise, within the condition where the source belongs to a group which is accurately assessed as likely to be knowledgeable (adults) then we expect the infant to find incorrect labeling events more surprising. However, the strength of these effects will be mediated by the infant's expectation that their perception of the speaker is salient.

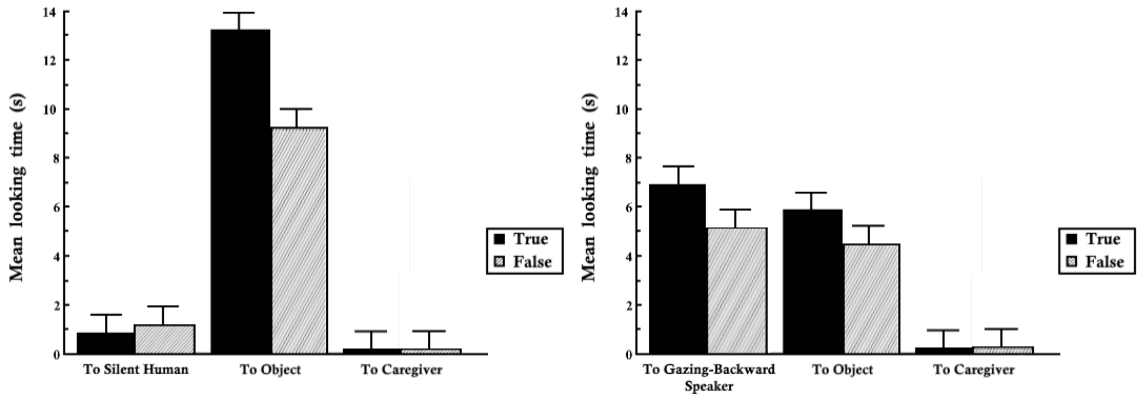
Figure 5.1: Looking time to parent, label source and label target for true and false labels from human (experiment one, left) and audio speaker (experiment one, right)



In other words, the pattern of looking times corresponds with the infants having a high degree of confidence in their prior beliefs that speech from audio speakers is not likely to be as reliably informative as speech from adult humans. Koenig and Echols (2003) also included a third experiment to ensure that the difference in infant responses was a result of infants interpreting the human *as a labeling source*, and not just an effect of their presence. In the presence of a silent human experimenter, infants tended to look away from the object longer during false labeling event without choosing to attend to the silent human. These results demonstrate that the infants' looking behavior is influenced by perceptions which are specific to the speaker.

In a fourth experiment, the investigators presented the infants with a human labeler who was situated with her back to the visual display. Consistent with the prior expectation that human informants are more informative than audio speakers, even in the condition where the human labeler evidently lacked perceptual access to the pictured objects, infants still preferentially directed their looking behavior towards her at a rate significantly higher than in either the audio speaker condition (figure 5.2, right) or the silent human condition (figure 5.2, left). These results demonstrate that non-linguistic speaker perception influences preverbal infants' behavior in response to labels (Koenig and Echols, 2003).

Figure 5.2: Looking time for true and false labels from audio speaker in presence of silent human (left) and from human gazing backwards (right)



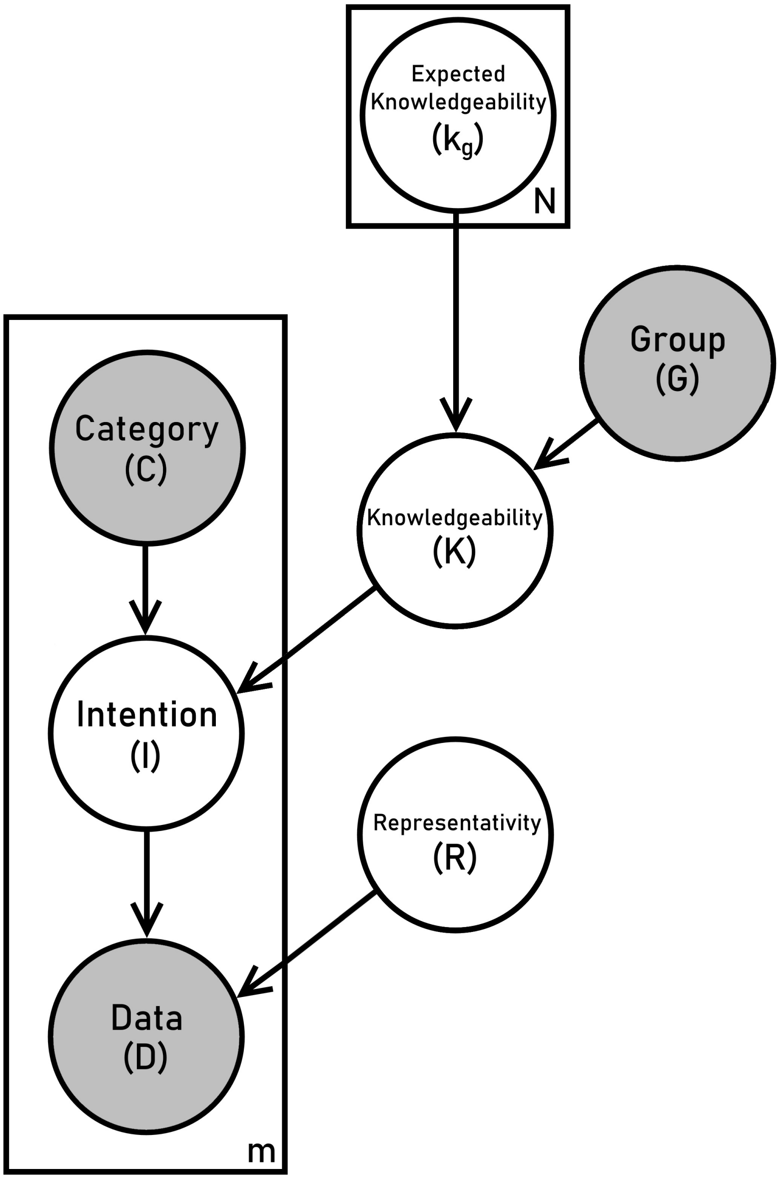
5.4.1 Modeling the knowledgeability of informants from different groups

In this section we will demonstrate how infant's beliefs about groups of informants predict their looking behavior during labeling. We characterize the infants' speaker perception as an inference on the knowledgeability of the speaker (K). We will assume a uniform prior over knowledgeability (K), the group membership of the speaker (G) and the phonological category (C), and contrast expectations about the effect of speaker perception on infant looking behavior in two conditions: with and without the production of labels. We will show that the model predicts the pattern of results found in Koenig and Echols (2003), lending support to the hypothesis that interactions between speaker perception and epistemic trust shape word learning in preverbal infants. Further study is required to determine how cues to epistemic trust may vary between learners from different social groups.

In the condition with a backwards-gazing adult speaker, infants did not look longer to the object than the speaker, or have different looking behavior on true and false labeling trials. This behavior may be contrasted with the experiments where the human labeler had visual access to the labeled object. The presence of perceptual access cues may have influenced the infants' uncertainty in the label by making it less clear what objects the informants are talking about - this can be modeled as comparatively low uncertainty about the category, C.

The graphical model for the first three experiments (featuring an audio speaker, a human labeler, and a silent human with visual access accompanied by an audio speaker) is given in figure 6.1. For these three experiments, we model the the category (C) as known, corresponding with the infant's access to the visual object display and lack of inconsistent perceptual access cues. The learner is modeled as needing to infer the speaker's knowledgeability (K). In the fourth experiment, where the labeling source was an adult facing away from the object display, I will interpret their gaze as a conflicting perceptual access cue, indicating that the subject of the labeler's speech is unknown, and both the category (C) and knowledgeability (K) must be inferred.

Figure 5.3: Graphical model with group



Assuming uniform priors, where speech sources are independently expected to be knowledgeable or to be animate in identical proportions, we may suppose there are four different types of speech sources the infants may perceive. We begin by supposing that the infant's goal is to learn about object labels in the presence of group-specific variation in knowledgeability. Further, we will assume that their beliefs about G and k_g are known to be correct. There are N possible unambiguous beliefs the infant may hold about the affiliation of a speaker, each giving rise to distinct expectations of the amount of information they will provide, where N designates the number of values G may take. In this example, $N = 2$.

We have assumed a number of phonological categories $n = 3$. If the infants attend only to the category, and not to the informant's group membership or knowledgeability then they must distinguish only three types of outcomes - those which correspond to the n categories. However, when knowledgeability is considered, using our simple distribution over the two events $(K = 0)$ and $(K = 1)$, there will be twice as many kinds of outcomes, for a total of six. Each communicative event is characterized by a pairing of K and C.

Our assumption that C and K are distributed uniformly essentially makes the problem of inferring its value as difficult as possible. Predicting which informants are knowledgeable with no additional clues is equivalent to guessing the outcome of a coin flip from a fair coin. In the context of labeling, we expect human sources to be experts while audio speakers are less reliable. Let us suppose for example, that infants are biased to believe that adult informants are very likely knowledgeable $P(K = 1|G_{adult}) = k_{adult} = 0.8$, while for audio speakers they are less trusting $P(K = 1|G_{audio}) = k_{audio} = 0.15$. The normalized probability of observing speech sources of the four different types is given in the table below, table 5.1.

Table 5.1: probability of informant type

K, G	$P(K, G)$
knowledgeable adult	0.4
knowledgeable audio speaker	0.075
unknowledgeable adult	0.1
unknowledgeable audio speaker	0.425

In this example, adults would be 84% of knowledgeable sources, but only 19% of unknowledgeable sources. The broad trend that the infants look longer to adult speakers can be described by a preference for knowledgeable speakers coupled with the accurate belief that adults are more knowledgeable about labels than other types of informants. This pattern predicts the main effect found in Koenig and Echols (2003), that infants spend more time attending to the adults than the audio speakers, independent of manipulations to the perceptual access of the source or the accuracy they displayed. To demonstrate that this prediction follows from an expected prevalence of knowledgeability among human speakers, we will compare expectations about the speaker's knowledgeability for each type of speaker with and without speech, showing that the group which is more knowledgeable will also have more informative speech. For this we will use the Kullback-Leibler divergence.

Supposing that C and G are known to the infant, we can model the infant's behavior in the first three experiments as an inference about K, showing that infants' looking time to the object are predicted by their expectations about the informants' credibility. For informants who are associated with more certainty about K (the adults), infants look more to both the object and the labeler than for informants who are associated with less certainty about K (the audio speakers). Observing either an adult or an audio speaker for the same amount of time is expected to result in two different amounts of information, defining an abstract expectation of comparative information density. Infants looking longer to adults informants are performing in line with the expectation that adult speech is more informative than that of audio speakers.

5.4.2 The effect of affiliation on the value of speech

Supposing that a label is observed, the impact on the listener's beliefs will be different depending on their beliefs about the affiliation of the speaker. We model this with a measure of information, the Kullback-Leibler divergence, which describes how differently knowledgeability is expected to be distributed either with or without a phonetic observation D, given a category C and an affiliation with group G. Equation 5.1 below describes this measure.

$$D_{KL}(P(K_w|D_d, G_g, C_x)||P(K_w|G_g, C_x)) = \sum_w P(K_w|D_d, G_g, C_x) \log \frac{P(K_w|D_d, G_g, C_x)}{P(K_w|G_g, C_x)} \tag{5.1}$$

Table 5.2 describes the relative entropy (KL divergences between prior and posterior probability) of the speaker's knowledgeability given four hypotheses about the circumstances of the labeling utterance - that the speech is either unambiguously supportive (a true label) or contradictory (a false label) to the given category, and that the speaker is either an adult or an audio speaker. A derivation appears in section 7.2.

Table 5.2: KL divergence of K with D compared to without for each type of observation

condition	$\sum_w \log P(D_d \vert K_w, G_g, C_x) P(K_w \vert G_g, C_x)$
correct adult	0.0846
incorrect adult	2.3219
correct audio speaker	0.1701
incorrect audio speaker	0.2345

The KL divergence is greatest for false labels issued by human speakers, consistent with the experimental results that infants looked longest to these types of labelers. The prior expectation that adults are more likely to be knowledgeable is biases the learner to regard overtly erroneous data from these speakers as the most informative, as reflected by the greatest relative entropy belonging to this type of observation. These predictions are consistent with the pattern of experimental results, which show significantly longer looking times to human informants than to audio speakers, and longer looking times to incorrect human informants than to correct ones.

Expectations about the epistemic reliability of sources conditioned on beliefs about their affiliation predict different expectations for both the information density of an informant's signal, and the resilience of their signal to error. For example, an adult is expected to provide more information than the average informant, while the average informant is expected to provide more information than an audio speaker, effectively defining a hierarchy of information density among informant types. Secondly, the model predicts that the amount of information provided by an adult who has been mistaken for an audio speaker is greater than vice versa. Although the present example strains credulity, I will return to a discussion of ambiguity and perceptual error in group membership and its effect on expectations of informativity in chapter 6.

Abstractly, the infant in our model identifies adults as more reliable informants and assesses these identifications as accurate. This does not imply that adult informants are perceived to represent more information about the category itself, rather that they more consistently provide evidence that their testimony about the category can be trusted to predict the performance of other speakers.

In essence, speakers who are expected to be knowledgeable are also expected to act alike, making them more reliable sources of information for K. In labeling an object, a knowledgeable speaker provides evidence for that label, but also against other candidate labels. Here we can interpret the low information expectation when the speaker is unknowledgeable as a belief that the speaker is not especially likely to provide compelling evidence either for or against the true label, but is instead equally likely to proffer evidence supporting any hypothesis.

The incorrect adult informant described in the second row of table 7.22 is expected to provide far more information about knowledgeability than any other type of speaker.

This measure again suggests that even without a bias to perceive adults as more likely to be speech agents than audio speakers, the belief that they are likely to be knowledgeable can explain infants' predisposition to attend to these informants over others.

When we contrast the infants' responses to informants who are associated with certainty about both K and C (the adults with perceptual access), and those who are not (the audio speakers and backwards-gazing adult), we find evidence that the proportion of infants' looking time to the object and the labeler may be influenced first by the expectation of reliability and second by the perception of reliability in the labels themselves.

The prior belief that an informant belongs to a group shapes expectations of the information needed to encode variables which describe their speaker qualities. This effectively predicts the task complexity of making discriminatory judgments between groups, within groups and under uncertainty about group membership, prior to observing speech. The infants' expectation that their own perception be veridical predicts that they will prefer to attend to the adults over the audio speakers in general and specifically for false labels. In other words, infants expect that speakers who they believe to be adults will behave consistently with the expectations they have for adults, while speakers who they believe to be audio speakers will more often defy the expectations they have for audio speakers. This is a rather elaborate way of saying that if the infant knows adults are more informative than audio speakers, they will also know that differences between adults are more informative than differences between audio speakers.

5.4.3 Discussion

While traditionally, models of language acquisition have assumed that early lexical knowledge is acquired independent of social cognition, this model places preverbal linguistic learning within a framework of social knowledge. In this series of simulations I have used judgements of epistemic trust to operationalize social judgements about language varieties.

The affiliative model of word learning presented here predicts that the pattern of infants' looking to objects may vary according to two factors - beliefs about the category being labeled, and about the reliability of the speaker. In this experiment, the objects and labels were familiar to the participants, making it equally easy for the infant to infer the category regardless of their prior beliefs about speakers. This is consistent with the results of the original experiment, which showed no significant difference in looking to the object between speaker types.

Contrastively, the pattern of infants' looking to informants reflects a preference for adults, the group we have assumed they believe to be more knowledgeable. I argue that inferring the knowledgeability of the informant fundamentally changes the listener's expectations of the informant's value, and therefore how much the learner expects to benefit from attending to the informant for both non-linguistic and linguistic information. In effect, longer looking times to adults evince the belief that adults are epistemically more trustworthy, and are consistent with the model predictions. Further, owing to their relative infrequency, seeing an adult name an object with an incorrect label is more informative about their knowledgeability than seeing them name the object with a correct one. This means that despite being frequently incorrect in labeling, unknowledgeable speakers are predicted to elicit more attention than knowledgeable ones. Although the data that they provide is expected to be largely inaccurate with respect to the category, it is expected to be largely accurate with respect to identifying other inaccurate speakers.

The experimental results are consistent with a story wherein the infants effectively demonstrate comprehension of two presented language varieties, with a preference to acquire of only one “dialect:” that spoken by the adults.

Acquiring a group-specific dialect would require the learner attend to information which distinguishes both linguistic variation within this group, and the non-linguistic variation which distinguishes members of that group from others. The pattern of results in Koenig and Echols (2003) is consistent with this explanation of infants’ preferentially attending to informants according to rational beliefs about the information density required to represent their behavior, including but not limited to speech.

Acquisition of the standard dialect, as we have defined it, would not require accurate discrimination of informants from different groups. The model introduced here accounts for existing findings regarding infant responses to familiar and novel labels from humans and audio speakers as a function of epistemic trust. The model makes further testable predictions about how differences in beliefs about the distribution of knowledgeability within and between populations, and beliefs about the accuracy of those beliefs produce differences in task complexity for different learners. Listeners who are acquiring different combinations of dialects can be predicted to perform differently on tests of both the recognition of familiar labels and the acquisition of novel ones.

In the previous chapter, I suggest that the existing literature is compatible with the hypothesis that infants use inferences about knowledgeability to guide their acquisition of novel words. In this chapter I have shown that this assumption also predicts how infant responses to true and false labels will vary as a function of informant type, arguing that infant behavioral expectations are conditioned on categorical judgements about the informant. In the model, the infant's evaluation of the accuracy of their own judgements is conditioned on beliefs about groups, and beliefs about the accuracy of those beliefs. Affiliative judgments provide predictions about both the potential informational content of an informant's behavior and the accuracy with which this content can be perceived. These findings suggest that even preverbal infants monitor their own cognition and are selective about the kinds of sources they learn words from.

Epistemic trust evidently informs the interpretation of lexical knowledge, and may influence how new words are acquired earlier than previously supposed. If preverbal infants have beliefs about their own membership in socially defined groups, we should predict significant differences in performance on tests of word learning by children with distinct social backgrounds. The function of attentional and memory processes controlling lexical-phonetic perception in learners is therefore predicted to vary even among speakers who share a target dialect, by virtue of the distinct social beliefs entailed by membership in different groups.

This model also has the descriptive power to capture listener beliefs about labeling behavior which span variation across dialects. Demonstrating the belief that two instances of a label refer to the same referent independent of accent requires infants to control for any impact of their beliefs about the accented speech on their interpretation of its referential content. Beliefs about the characteristic accuracy of informants owing to their membership in either linguistically or non-linguistically defined groups will necessarily exert an a priori influence on infants perception on both the content and reliability of their speech behavior.

It is critical for studies of language development to account for how speaker perception varies across populations. The findings suggest further study is needed to differentiate how infant attention to speakers and their labels differ between cultural groups, and the effect of biases in speaker preference on performance in tasks of lexical comprehension and word learning. Insofar as different populations of learners may be predisposed to different attitudes towards speakers, prior beliefs about groups may significantly affect task difficulty.

In opposition to previous work, the present model defines a linguistic basis for the perception of differences between standard and non-standard language varieties. Although it is generally assumed that these distinctions are not linguistic in nature, there is not yet clear evidence upon which to draw a conclusion, as the subject has largely not been investigated (Milroy and Milroy, 2012). In modern linguistics, the concept of a standard dialect has generally been defined as having no structural distinction from non-standard dialects. Standard and non-standard dialects are generally understood to be distinguished in a purely non-linguistic social dimension, and it is assumed that there is no linguistic basis for discriminating between standard and non-standard language varieties as such. By assuming that expectations of standard and non-standard language may be defined as differing beliefs about the distribution of knowledgeable informants, the present model can describe the acquisition problem for learners who are acquiring a standard dialect as distinct from the problem facing learners who are acquiring multiple constituent dialects in addition to a standard variety.

Early emerging preferences for native speakers demand that studies of language processing carefully control for the perception of the speaker's dialect. However, many researchers in psycholinguistics assume that presenting experiment participants with a single voice is sufficient to control for bias introduced by variation in the presentation of indexical features. The model defined here predicts that differences in perception of both linguistic and non-linguistic cues in the speech signal are expected to impact infant perception of labels. To wit, this model successfully explains the looking behavior of preverbal infants in response to true and false labels from sources affiliated with groups that are more (adults) or less (audio speakers) trustworthy.

In the next section I will summarize this dissertation, and explore its implications for studies of language development, especially with respect to measures of online language processing and vocabulary size. I will argue that differences in vocabulary size reflect intrinsic social stratification in a population of language users.

Insofar as vocabulary varies within populations, so do social identities. Measures which predict vocabulary are predicted to correlate with both measures of SES and of cognitive control. Educational interventions which seek to encourage in disadvantaged populations the development of linguistic skills similar to those of high SES language learners must be sensitive to the pre-existing social knowledge and potential social motivations of young learners, as well as possible early emerging phonetic processing biases for affiliative cues associated with varieties of their native language. The pattern of results suggests a role for phonetic contrasts which are not strictly referential in guiding listeners' attention to labels.

Chapter 6

Summary and Next Directions

The goal of this work has been to investigate what effects the processes that infants use to categorize informants have on the phonetic perception and lexical acquisition of preverbal infants. What beliefs do infants form about the relative quality of speech sources, and how do these beliefs shape infants' attention to the input?

I began by describing the evidence that judgements of informant quality in both non-linguistic and linguistic domains appear to rely on the same processes of selective trust, and detailing the evidence that infants attend to abstract speaker qualities, including membership in groups.

I then introduced a model describing the behavior of listeners in a label learning task. This model, adapted from Shafto et al. (2012), was originally applied to simulate the behavior of preschool aged children. However, I demonstrated that this model also effectively describes the behavior of 14-month old infants on a word learning task. I adapted this probabilistic model to account for 14-month olds' responses to labeling tasks of different accuracy arising from prior beliefs about the accuracy of informants according to their membership in groups.

Both of these simulations suggest that preverbal infants are epistemically evaluating sources of linguistic information. In effect, infants failing to show sensitivity to phonetic detail may be demonstrating an expectation that the source and/or their data are not trustworthy. Experiments which control for the infants perception of source reliability are needed to provide more explicit support for this interpretation of the literature.

Further, this work underscores the need for a research program exploring how infants form durable beliefs about linguistically defined groups. While phonological processing and the interpretation of non-referential speech features are often assumed to develop as distinct processes, to the extent that representational content from either domain may affect the evaluation of the source as epistemically trustworthy, the impact of each on early word learning cannot be dissociated without further evidence from experimental studies.

I argue that as perception of non-linguistically defined groups and perceptual access cues both influence learners' attention to linguistic cues and their beliefs about them, it is necessary to better define the role of social perception in language development. Psycholinguistic investigations must incorporate more sophisticated control measures to evince the absence of bias. Models of language acquisition which rely on the abstraction that infants wholly dissociate their perception of the speech from that of the speaker cannot predict the effects of source evaluation on acquisition. The model set forth in this dissertation provides an explanation for many effects already established in the literature.

In this chapter, I will discuss some potential ramifications of this model for linguistic study, and outline the potential for this model to provide a unified understanding of dialect perception and the time course of linguistic prediction and processing. Differences in beliefs about languages are expected to emerge early alongside differences in language background, predicting significant consequences for between-population measures of linguistic competency.

6.1 Informant preference and early linguistic development

"Language ideology" is a term which describes how listeners contextualize linguistic structures within a framework of cultural beliefs about social and linguistic relationships (J Irvine). Although social perception has been historically neglected in the study of early language acquisition, I argue that the early emergence of selective trust may lay the groundwork for sociophonetic attentional preferences which will shape the formation of early language ideology and therefore the lexicon.

Existing models of language acquisition largely rely on the abstraction that all speakers use the same language variety, however this assumption that infant speech perception is dissociable from speaker perception has become an impediment to understanding how acquisition of all manner of linguistic phenomena is achieved. Contrastively, the assumption that even young infants may have principled beliefs about the impact of speaker identity and social status on the meaning of speech provides us with the basis to contextualize their linguistic development within the constraints implied by their developing social knowledge.

Assuming the abstract assumption by the learner that within a language there are at least two discernible varieties corresponding to the native (knowledgeable) dialect and non-native (unknowledgeable) dialects yields many interesting kinds of predictions. First, I will describe two ways differences in the perception of informants are expected to impact the quality of listener speech representations: effects of speaker appearance on linguistic beliefs, and effects of developing learning biases on linguistic development trajectories. Second, I will outline the predictions this model makes for the impact of speaker perception on the time course of linguistic processing, how these may change depending on the number of dialects the learner is acquiring, and the implication for listener judgements of salience and appropriateness. Lastly, I will comment on how this scientific inquiry is socially situated.

6.2 Learning about speech from non-linguistic features

Insofar as perception of social group membership is implicated in infant speech processing, this model implies that listener interpretation of non-linguistic cues as markers of affiliation will necessarily mediate beliefs about knowledgeability. These cues may range from facial features or expressions to phonetic variation. In this section I will introduce an analogy to describe how affiliative features and environmental cues may be recruited to guide word learning.

Suppose you are shopping at a big-box store. You notice that the signage, the walls, and the shelving all feature the color red. You want to know which aisle you should visit to find a blicket. You can see two people nearby. How might you identify which of the two people is more likely to be knowledgeable?

Let us suppose that one of the two people is wearing a red shirt. You may intuit that this feature is indicative of being employed at the store with the red decor. In this context, the people in the environment can be broadly categorized as belonging to a latent group describing their status as employees or customers. In the context of the store, we expect employees to be experts, so we will define knowledgeability as dependent on group. Let us suppose that employees are overwhelmingly knowledgeable $P(K = 1|G = 1) = 0.95$, and customers less so $P(K = 1|G = 0) = 0.1$.

We also expect employees to share some observable features which distinguish them from customers. In this example, the salient feature is shirt color. We define a variable F to encode whether a specific person is observed to be wearing a red shirt. Assuming the incidence of red shirts is dependent on the latent feature of employment, G, we suppose that only one out of four customers are expected be wearing red shirts at a given time $P(F = 1|G = 0) = 0.25$, while almost all employees are wearing them $P(F = 1|G = 1) = 0.9$. The four possible combinations of features are given below in table 6.1.

Figure 6.1: Graphical model with group, feature

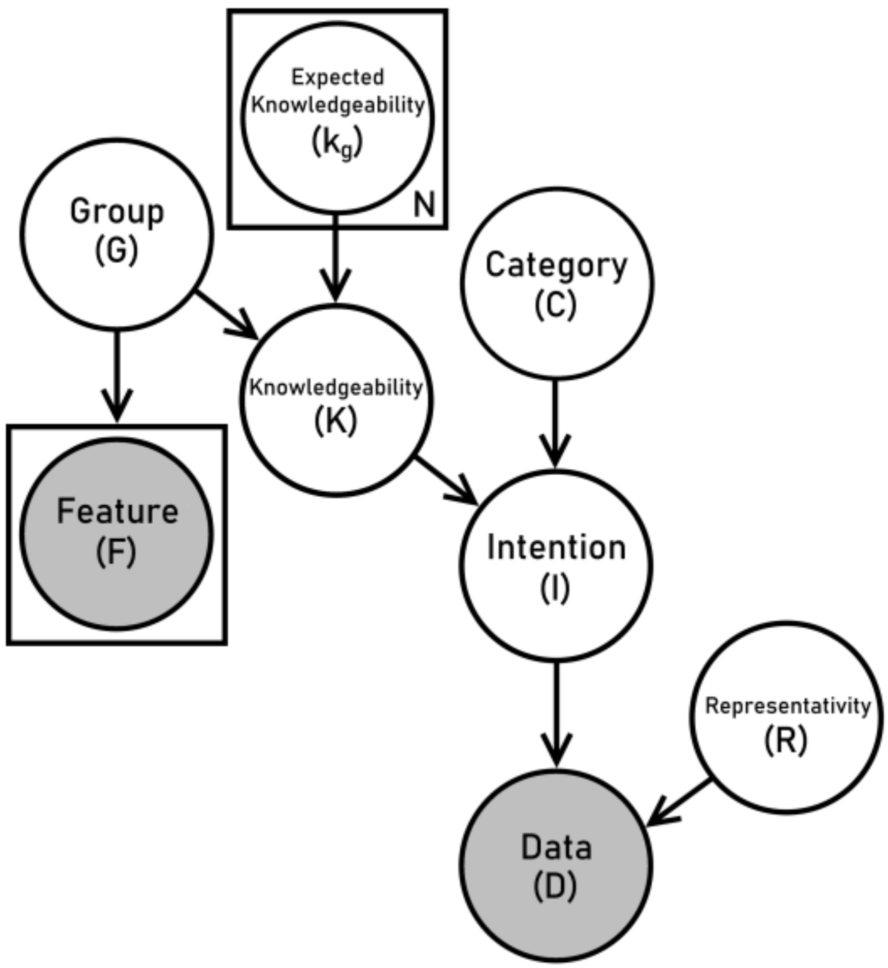


Table 6.1: Incidence of speaker characteristics

G, F_1	G	F_1	$P(K=1\|G)$	$P(G, F_1)$
employee with red shirt	1	1	0.95	0.90
customer with red shirt	1	0	0.10	0.25
employee without red shirt	0	1	0.95	0.10
customer without red shirt	0	0	0.10	0.75

The probability that a person is knowledgeable given that we have observed their shirt to be red is the sum of the probability that they are a knowledgeable employee with a red shirt, or a knowledgeable customer with a red shirt. The posterior probability on knowledgeability after observing whether the informant has a red shirt is given in equation 7.7 below.

$$P(K=1|F) = \sum_G \frac{P(K|G)P(F|G)P(G)}{\sum_{G'} P(F_1|G')P(G')} \tag{6.1}$$

We assume that knowledgeability, K, is independent of F, when G is known. That is, while wearing a red shirt (F) provides evidence of employment (G), if we know that someone is an employee, the color of their shirt does not give us additional information about their knowledgeability. This may be expressed as an information gain of zero bits after observing F.

Let us suppose that only one out of four customers are expected to be displaying feature F_1 (wearing red shirts) at a given time $P(F_1 = 1|G = 0) = 0.25$, while almost all employees are wearing them $P(F_1 = 1|G = 1) = 0.9$. Supposing that 20% of the people in the store are employees P(G = 1) = 0.2, the probability of observing each possible type of person is given in table 6.2.

Table 6.2: Types of people in the store conditioned on shirt color

G, F_1, K	G	F_1	K	$P(K, G\|F_1)$
knowledgeable employee with red shirt	1	1	1	0.450
knowledgeable customer with red shirt	0	1	1	0.053
unknowledgeable employee with red shirt	1	1	0	0.024
unknowledgeable customer with red shirt	0	1	0	0.474
knowledgeable employee without red shirt	1	0	1	0.031
knowledgeable customer without red shirt	0	0	1	0.097
unknowledgeable employee without red shirt	1	0	0	0.002
unknowledgeable customer without red shirt	0	0	0	0.871

Given that we have observed a person wearing a red shirt, we can then calculate the probability that they are knowledgeable by summing the probability that they are a knowledgeable customer or a knowledgeable employee. The probability of a speaker being knowledgeable given an observation of feature F_1 is given in table 6.3. The probability of our red-shirted informant being knowledgeable is (0.503), roughly four times the posterior probability on knowledgeability for the non-red-shirted informant (0.127).

Table 6.3: Inferring knowledgeability from shirt color

F_1(red shirt)	$P(K=1\|F_1)$	$P(K=0\|F_1)$
1	0.503	0.497
0	0.127	0.873

If certain *kinds* of informants are more likely to consistently provide useful information (e.g. store employees), then those features which tend to distinguish these informants from other kinds (e.g. shirt color) are distributed along a hierarchy of informativity about the subject of knowledgeability (e.g. the location of the blickets or the correct label for a dog). Suppose, for example, we instead attended to a different observable feature of our potential informants, (e.g. hair color). If we observe feature F_2, whether the informant has red hair, and assume hair color is distributed independently of employment status, we may express the probability of observing this feature F_2 with $P(F_2|G) = P(F_2)$.

In this case, the feature is uninformative with respect to the latent variable of employment, and therefore with respect to knowledgeability. The probability of knowledgeability given observation of this feature is given in table 6.5. The independence of G and F_2 is apparent in that each row in the table is identical. In other words, the posterior probability on knowledgeability is the same for all values of F_2.

Table 6.4: Inferring knowledgeability from hair color

F_2(red hair)	$P(K=1\|F_2=1)$	$P(K=0\|F_2=0)$
1	0.27	0.73
0	0.27	0.73

A rational learner given a choice of cues will preferentially attend to informants displaying features which are most informative about their knowledgeability K and latent group membership G. In other words, if we want to learn who knows the location of the blickets, we must pay more attention to shirt color than to hair color. Equation 6.2 describes how the numbers in table 6.5 are combined to calculate the information gain, shown in table 6.6, which describes the expected reduction in uncertainty on the knowledgeability of the informant from observing either shirt or hair color given knowledge that the informant is either an employee or a customer. As per our assumptions, the observation of hair color is equally uninformative in distinguishing knowledgeable employees from customers, but shirt color is significantly more useful for distinguishing knowledgeable speakers from unknowledgeable speakers among employees compared to among customers.

$$I(K;F|G) = H(K|G) + H(F|G) - H(F,K|G) \tag{6.2}$$

Table 6.5: Information about knowledgeability by group

G	$H(K\|G)$	$H(F_1\|G)$	$H(F_1, K\|G)$	$H(F_2\|G)$	$H(F_2, K\|G)$
employee	0.1582	-0.6356	-0.1670	0	0.1582
customer	0.1547	-0.0549	-0.7560	0	0.1547

Table 6.6: Inferring knowledgeability from shirt and or hair color

G	$I(K; F_1\|G)$	$I(K; F_2\|G)$	$I(K; F_1, F_2\|G)$
employee	0.6268	0	0.6268
customer	0.8558	0	0.8558

To return to the language problem - supposing that certain kinds of linguistic informants are more likely to consistently provide useful information (e.g. store employees), a learner who wishes to identify those sources which best serve as linguistic models is best served by attending to those features which distinguish them from other kinds of informants (e.g. shirt color). The higher expected information gain on knowledgeability for customers shown in table 6.6 demonstrates that shirt color is more useful for distinguishing knowledgeable and unknowledgeable customers than it is for distinguishing knowledgeable and unknowledgeable employees.

Supposing that infant beliefs about groups must arise from are attending to and generalizing their learning from features which categorize informants in non-linguistic ways, what effect do these features have on the evaluation of linguistic informants? On other words, how do infant beliefs about the recognizability of speakers and their membership in social groups shape attention to and beliefs about the linguistic behavior of different kinds of native language speakers?

David J Kelly et al. (2007) demonstrates that by 9 mo infants show perceptual narrowing in their facial recognition, more successfully discriminating faces of own-race people than those of other races. If infants' speech expectations are driven by perception of speaker affiliations, such as race, we should expect infants to interpret phonetic information differently depending on the race of the speaker. Weatherhead and White (2018) provides some preliminary experimental evidence to support this claim. In this study, 16 month old infants were exposed to familiar words either in a familiar or unfamiliar accent. The findings suggest that at 16 months, infants possess the expectation that familiar-race speakers are likely to pronounce words in familiar ways. Further, the infants did not have this expectation of unfamiliar-race speakers, suggesting that the infants are using knowledge of epistemically trusted groups to represent the native accent and guide interpretation of variation in labels.

Under the account presented in this thesis, preferences for various cues associated with social categories will affect the linguistic behavior of the listener by exerting an influence on their beliefs about the informant's production and perception patterns *before they have spoken*. The expectation that a speaker belongs to a familiar group suggests he will also behave in a familiar way. The expectation that the speaker behaves in a familiar way also suggests the speaker is recognizable as a member of a familiar group. Both perception of the speaker's group membership and their expected reliability conditioned on this perception are expected to predict listener behavior in response to familiar and novel labels.

Infants may also therefore be sensitive to the phonetic patterns which distinguish linguistic varieties associated with the socioeconomic affiliation of their caregivers. However the research in the infant speech perception is overwhelmingly biased to privilege description of the developmental trajectory seen in the language skills of middle and high socioeconomic status (SES) learners.

Kathleen H Corriveau, Kurkul, and Arunachalam (2016) exposed children of varying SES backgrounds to informants who either used passive or active voice constructions to describe images. It was found that children from high SES backgrounds show a preference to learn novel words from informants who used the more complex syntax, while children from lower SES backgrounds preferred the informants who used simpler sentence structure. The finding suggests that despite both groups of children demonstrating understanding of the more complex syntactic form, that the relative amount of experience with informants who use each form predicted their selective trust in novel informants exhibiting the same behavior (Kathleen H Corriveau et al., 2016).

Positing that there is one language variety which is considered more representative of knowledgeability than others predicts that a speaker who recognizes multiple dialects of a language will nonetheless favor one dialect over others. Such a model can potentially begin to account for how learners develop principled beliefs about the content of speech from same race and other race informants, as well as the role of vernacular and standard dialectal items and structures within a given community of practice. Early in development we expect listeners to respond differently to dialects not solely as a function of exposure, but also as a function of attitudes towards the speech determined by the quality of that exposure.

6.2.1 Accuracy of affiliative perception impacts informativity

Knowing about groups abstractly allows the listener to make metacognitive predictions about their perception of the speaker's behavior which are predicated on the accuracy of the affiliative perception. Supposing that the listener may make perceptual errors in interpreting affiliation, the presence of an increasing number of groups would yield an exponentially increasing number of potential types of errors in affiliative perception. However, the belief that some groups rightly appear to be relatively more or less informative than an abstractly defined standard provides the basis for cognitive strategies which selectively license attention to affiliation by distinguishing two types of affiliative errors: believing that an informant belongs to a group that is more knowledgeable than the group the informant truly belongs to, and believing that the informant belongs to a group which is less knowledgeable than their true group. Different beliefs about the group membership of a particular informant are therefore correlated with distinct beliefs about the likelihood of these errors.

Both of the experiments modeled in this dissertation examine typically developing infants, and rest on the assumption that their beliefs about the distribution of the informants are accurate. However, we may also potentially extend this model to describe some atypically developing groups as relying on distinct beliefs about the distribution of knowledgeability. Differences in perceptual sensitivity to epistemic cues are expected to cause cascading effects for language development. For example, tv watching is correlated with language delays (Chonchaiya and Pruksananonda, 2008), suggesting that exposure to televised speech models may be one factor retarding the development of sensitivity to epistemic cues which distinguish the performance of knowledgeable informants. Similarly, children with Autism Spectrum Disorder do not reliably follow gaze (Carpenter et al., 2002), and can be expected to have different beliefs about informants based on different perceptual strategies. Children with ASD appear more likely to determine the referent of adult speech using their own gaze, as opposed to the gaze of the adult speaker (Baron-Cohen, Baldwin, and Crowson, 1997). Although this demonstrates a failure to utilize referential intent in a typical way, autistic children have nonetheless been shown to perform comparably to typically developing children in mapping novel words to referents under conditions of referential ambiguity (Preissler and Carey, 2005). The language delays observed in ASD children could be the result of atypical integration of social and linguistic cues. There is further evidence that autistic children and children who are at-risk of diagnosis with ASD do not correctly interpret the underlying communicative intention of gaze (Gliga et al., 2012). These results support the theory that this population has distinct beliefs about the distribution of knowledgeability compared to typically developing children.

The conjecture that either mono-dialectal and multi-dialectal TD children must learn to discriminate between linguistic groups with different characteristic knowledgabilities allows us to make predictions about functional differences in these populations as well. Inaccurately perceiving the affiliation of a speaker would necessarily cause the learner to systematically either overestimate or underestimate the likelihood that the speaker is knowledgeable, in accordance with variation between the expectations for each informant type. Supposing that the infant has already learned to accurately identify members of different groups and broadly characterize their behavior allows us to further assume that their perception of informant affiliation is minimally errorful and unbiased. Despite assuming that the learner lacks any perceptual biases which cause them to make *unfair* errors, we still expect them to express epistemically based beliefs which impute higher informativity to some groups than to others. That is to say that perceiving a customer to be an employee would result in an over-estimation of the informant's knowledgeability relative to a constructed standard, while perceiving an employee to be a customer describes an underestimate relative to that same standard.

When group membership is not taken into account, then the observer may effectively employ a single standard in perception, having identical expectations about information in K and C for all observed informants. In this case all informants are expected to provide the same quality of speech. However, when the affiliation of the informant is known, then their membership in a group provides the basis for a distinct behavioral standard - for example, adults are expected to provide more information about K than audio speakers. Consequently, sampling C and K jointly conditioned on any value G is expected to yield less information than sampling these two variables without conditioning. In effect, this describes an expectation about the relative value of each type of informant to the listener's goal of perceptual accuracy.

The model predicts that the average informant should be expected to be more reliable than audio speakers, and less reliable than adult human speakers. Knowing that the informant is an adult then allows the listener to make metacognitive predictions about the accuracy of their perception of the speaker's behavior, given the assumption that this affiliative judgement is correct. Metacognitive predictions about the accuracy of speaker perception are necessarily predicated on the accuracy of the affiliative perception. Although the simulations presented here rely on the assumption that metacognitive predictions are maximally uninteresting, further study is required to tease apart under what circumstances this may or may not be true.

Conceptually, this model predicts that successful mono-dialectal learning requires appropriately avoiding attendance to speakers' membership in linguistically defined groups. However, learners who are acquiring additional dialects must attend to information about social groups and acquire distinct representations of dialects as social group markers. As a result these learners may perceive additional ambiguity in lexical items - in effect, for the multi-dialectal learner, inferences about knowledgeability are constrained by additional beliefs about dialect groups that mono-dialectal speakers lack. Multi-dialectal learners are accordingly predicted to display decreased ambiguity in speaker perception. I argue that measures of language ability which do not contrast ability in both perception of the speaker and the spoken are implicitly biased to privilege children of specific cultural backgrounds. The requirement that the listener manage ambiguity between dialects implicates cognitive control, providing a rich framework in which to make predictions about individual differences in psycholinguistic performance both between and within-groups.

6.3 Language standards and culturally based beliefs about informants

The "standard" variety of a language is a variety frequently judged as more correct and acceptable than dialectal varieties. It is considered more prestigious than other varieties of the language, and associated with education, mainstream media, government and industry. Owing to the pervasive institutionalized nature of the standard language variety, speakers of a language are generally assumed to be passively proficient in this variety regardless of which dialectal varieties they produce. But how do such sociolinguistic judgements emerge?

Typically developing children acquiring linguistic competence also acquire culturally determined social perceptions and therefore acquire a culturally determined typical sociolinguistic competence. However, the relationship between linguistic and non-linguistic cognition is poorly understood. The assumption that early language acquisition is not subject to sociolinguistic factors is attributable to two main factors: first, it is part of a history of linguistic discrimination, whereby the language varieties of socially disadvantaged speakers are institutionally devalued and treated with bias by the language science community. This is true of racialized dialects (Baugh, 1988), Deaf language (Lane, 1992) and otherwise stigmatized dialects such as Southern English (Oetting, Lee, and Porter, 2013).

Secondly, the belief that early language acquisition is not influenced by social knowledge is supported by evidence that language acquisition outcomes are heavily influenced by the degree of exposure to language. This licenses the inference that differences in young listeners' attention to different language varieties are influenced most strongly by differences in the amount of exposure. In abstracting away non-linguistic variation between sources and treating infant speech perception as developing independently from social perception, traditional models of early language acquisition adopt both of these assumptions (Nardy, Chevrot, and Barbu, 2013).

However there is relatively scant research on the subject and it is conflicting; the sheer diversity of linguistic modalities, registers and non-linguistic indexical cues creates a staggering diversity of potential sociolinguistic beliefs to investigate within any given narrowly defined linguistic community. Broadening the scope of investigated sociolinguistic knowledge to include cross cultural phenomena like beliefs about language standards introduces even more potential complexity. This state of affairs calls for a unifying framework to guide the investigation of whether social knowledge is a factor in infant language acquisition.

The affiliative model of language acquisition presented in this dissertation provides a simpler explanation, defining both linguistic and non-linguistic social knowledge as dependent on beliefs about the reliability of informants. This theory asserts that learner attention to language varieties likely varies as a result of not just exposure but listener attitudes. This model has the power to predict a number of sociolinguistic phenomenon in addition to accounting for the acquisition of linguistic knowledge not traditionally considered to be sociolinguistic. For example, the model predicts that learners will develop divergent sociolinguistic beliefs even with identical exposure, supposing that they have different prior beliefs about social affiliation.

The model also predicts that learners may have beliefs about congruity between the speaker and the spoken. In previous sections I have theorized that perception of multiple social groups, such as the examples given of adults and audio speakers, or of customers and employees, effectively defines the perception of multiple group specific dialects, which may be either authentically or inauthentically performed. An informant may either be accurately expected to behave like a customer or accurately expected to behave like an employee. Likewise, allowing for inaccuracy in affiliative perception, informants may also be *inaccurately* expected to behave like a customer or *inaccurately* expected to behave like an employee. In addition to accurately interpreting the referential content of speech, learners must also acquire the ability to judge speech as congruent with some speaker identities and not with others. For example, labeling an item with "perro" may have the same referential function as "dog," but each will be judged as more congruent with some speaker identities than with others (e.g Spanish vs. English language users). In both cases the labeling action may be accurately identified (indicating a dog) independent of the speaker perception (believed to speak Spanish or believed to speak English). However, the listener is nonetheless expected to be sensitive to mismatches between their linguistic speaker perception and the observed word form (e.g. expected to speak English but producing "perro").

The learner's inclination to preferentially jointly attend to a speaker as a source of both negative and positive evidence is expected to vary as a function of the listener's beliefs about the speaker. In effect, we expect speakers who are believed to be likely knowledgeable to command attention to their speech simultaneously as both positive and negative evidence. Contrastively, a speaker who is perfectly unknowledgeable is not expected to demonstrate a relationship between the time course of these two events. Learners with different beliefs about the characteristic knowledgeability of groups are expected to consequently hold different preferences for language varieties, and perceive identical tasks to have different complexity.

Existing studies of online language comprehension in infants and children frequently make the assumption that infants are not sensitive to affiliative cues in the presentation of speech items. In this section I outline how the model introduced in this dissertation makes principled predictions about how the availability of referential information (the category

C, which refers to the object) and indexical information (the knowledgeability K, and representativity R) varies with respect to confidence in affiliative cues.

6.3.1 Time course availability of referential information

Traditional models of language acquisition which do not include non-linguistic social information therefore also do not predict a role for this information in the online processing of speech. However, the assumption that speech must be differentiated between dialects implies a central role for disambiguation in all levels of speech processing, and predicts differences in the time course availability of referential content and perception of iconic speech forms when hearing an utterance, according to the listener's pre-existing biases. The affiliative model of language acquisition makes principled predictions about how differences in beliefs about and preferences for certain social groups are expected to result in different time courses for both speech recognition and lexical acquisition.

For example, mono-dialectal speakers are predicted to perceive decreased phonetic ambiguity compared to their multi-dialectal peers, owing to a more simplistic representation of socially marked language variation. Mono-dialectal speakers inferences about object labels are predicted to be faster and rely on more precise representations as a consequence of these speakers perceiving fewer linguistically distinct social groups. Conversely, owing to the greater complexity of their social representations, multi-dialectal speakers are predicted to have more diffuse representations and slower reaction times but greater sensitivity to affiliative cues. Existing evidence shows that infant sensitivity to talker variation depends on native language experience (Johnson, Westrek, Nazzi, and Cutler, 2011; Fecher and Johnson, 2018). Further experimental work is required to determine the usefulness of comparisons between mono- and multi-dialectal modes of perception.

The affiliative model predicts that not only will the learner's beliefs about the speaker's affiliation affect online speech processing, but also the learner's perception of the accuracy of those beliefs. We may model the listener's expectation of knowledgeability as a Beta distribution over coin flip biases. When the probability mass is symmetrically distributed around this mean, listeners are expected to make equal numbers of errors which incorrectly identify knowledgeable speakers as unknowledgeable (false negative litmus test) and which incorrectly identify unknowledgeable speakers as knowledgeable (false positive litmus test).

For example, let us consider a learner who believes without observing any social group affiliation that a given speaker is perfectly knowledgeable with a likelihood of 50%. The learner is therefore maximally uncertain about the knowledgeability of the speaker before the sample of speech at time $t = t_0$ and after the sample at time $t > t_0$. Additionally they are maximally uncertain about the certainty with which the speaker's referential content can be interpreted at any time t. However, in the presence of cues which affiliate the informant with some social group, uncertainty about both the knowledgeability of the speaker and interpretation of referential content in the speech will be affected. In other words, even though in this scenario the listener is unable to verify the quality of their predictions about the speaker's knowledgeability before or after observing speech without observing any cue to group affiliation, they will nonetheless have predictable beliefs about the time course availability and reliability of both positive and negative evidence in samples of speech based on beliefs about potential group affiliation. This prediction accords with existing literature on how the presence of affiliative cues impacts online speech processing in adults. In particular, there is a wealth of evidence that the N400, a well-established ERP component encoding semantic violation, may also be involved in social learning (Osterhout, Bersick, and McLaughlin, 1997; Lattner and Friederici, 2003; White, Crites Jr, Taylor, and Corral, 2009; Siyanova-Chanturia, Pesciarelli, and Cacciari, 2012; Hehman, Volpert, and Simons, 2013).

As the mean prior belief about the speaker's knowledgeability increases above a likelihood of 50%, the listener's belief about what proportion of information in the referential inference can be derived prior to observing the utterance will increase. An increased bias to believe speakers are knowledgeable predicts a greater incidence of false positives, distributed over a smaller amount of information and therefore a greater surprisal for these events (the listener expects the speaker to be knowledgeable and concludes that the speaker is surprisingly, unknowledgeable). Conversely, an increased bias to believe speakers are unknowledgeable predicts a greater incidence of false negatives - (the listener expects the speaker to be unknowledgeable but then concludes that surprisingly, they are knowledgeable).

Depending on whether the speaker is suspected to be knowledgeable or suspected to be unknowledgeable, my model suggests that the listener's expected certainty about the referential content of speech must necessarily predict their listening strategy. A listener will make different predictions about the time course of positive and negative evidence in future observations according to their beliefs about the frequency with which they will encounter each kind of speaker. In order for two different language learners to acquire the same dialect, this theory predicts it is necessary for them to not only have similar exposure, but similar pre-existing listening preferences. A research program in early sociolinguistic learning is required to connect what is known about how social perception influences online speech processing in adults with the literature on development.

6.3.2 Between-speaker distribution of referential information

We have given an uncontroversial definition of standard language, however researchers often operationalize this distinction, so that studies conducted in different localities effectively attribute the label of "standard" or "mainstream" to distinct regional dialects. For example, Hosoda, Stone-Romero, and Walter (2007) operationalizes the standard dialect of English as a Northern Californian regional dialect. Operationalizing a definition of the standard variety as a particular regional dialect belies the objectivity of any comparisons made to non-standard varieties. Inherent in the definition is institutional power, implying a social imbalance between varieties and their speakers. "Non-standard" dialects are definitionally associated with racial, ethnic and socioeconomic identities which lack institutional power. Evidence suggests that membership in such groups is more strongly associated with gaps in academic achievement. For example, a study of cohorts of low, medium, and high SES school children in the United States, the United Kingdom, Australia, and Canada found that the U.S. had the most pronounced achievement gaps between children in all three SES categories (Bradbury, Corak, Waldfogel, and Washbrook, 2015). The authors found that 60–70% of the gap in math and reading scores found in school children before the eighth grade was already present at school entry. The consistent presence of and variation in achievement gaps across countries at the time of school entry implicates differences in the children's language skills arising from SES.

Effectively, identities associated with achievement gaps are also those identities associated with a lack of institutional power and therefore non-standard language varieties.

I assert that in class conscious societies, the recognition of a standard language effectively codifies social stratification between dialectal groups, and perception of dialectal group as central to the language learning faculty. This predicts that language users whose varieties are judged more consistent with the privileged mainstream variety will be afforded greater class mobility (Guy, 1988). Speakers of stigmatized dialects may unfairly lose educational, occupational and housing opportunities owing to prejudice against their dialect and the social group it is associated with. The ideal of equality among linguistic varieties is a scientific abstraction, but not a sociopolitical reality (Baugh, 1988).

In studying the acquisition of a standard variety across different dialectal populations, it is essential to recognize how this variety and its speakers are socially situated with respect to the entire range of dialectal variation. Utterances may be ambiguous with respect to whether they were produced in a standard or non-standard variety. If children are attending to dialectal variation, then their different experience with dialects must necessarily affect their perception of ambiguous forms and structures. Although the simulations presented in this dissertation have dealt with unambiguous utterances from speakers of ambiguous affiliation, the affiliative model also makes predictions about the result of ambiguous affiliative cues on the perception of ambiguous utterances.

Linguistic theory has generally held that dialect awareness generally does not emerge until later in development, and studies of perception in school age children suggest that adult-like identification of dialects emerges slowly (Wagner, Clopper, and Pate, 2014). However, it is important to distinguish between the ability to explicitly discriminate dialects and the influence of subconscious knowledge about dialectal variation. Evidence which suggests that children cannot recognize dialectal variation does not preclude the possibility that they may still be induced to attend to this variation by contextual and social factors. Further study is needed to determine how social cues may induce socially defined expectations of language use in infancy, and how sensitivity to these specific cues develops in different populations.

Insofar as the vernacular dialects of learners from high socio-economic status backgrounds are expected to closely align with the form of the standard variety, these children must be recognized as completing a comparatively simpler acquisition task than children whose home dialects are more dissimilar to the standard. The greater divergence between the standard variety and language varieties associated with low socio-economic status groups indicates that the challenge of lexical learning may be measurably more complex in acoustic, phonetic and phonological dimensions for these populations.

Unfortunately, in most studies of language acquisition, infants' potential recruitment of paralinguistic features is assumed to be adequately controlled for with the presentation of a single voice. Such a paradigm cannot effectively control for potential biases introduced by affiliative cues in the stimulus. As a result, I believe studies which compare the learning of different dialectal populations fail to accurately situate the abilities of their participants with respect to the standard variety. For example, studies of African American Language (AAL) routinely contrast the progress of children who are alleged to be learning *either* the "standard" variety of English *or* an ethnically coded dialectal variant.

In order to sort the children into these binary categories one must disregard that dialectal language exists along a spectrum. For example, there is significant overlap between SAE and AAE. Among children who speak AAE, the use of marked dialect features may vary (Washington and Craig, 1994; Washington et al., 1998). In order to accurately contrast the dialectal knowledge of children, it is necessary to more precisely characterize the dialectal behavior of so-called "standard language" users.

"Standard language" users invariably have their own vernacular and regional dialect features. Characterizing these speakers as only acquiring a standard variety conceals the typological distinctions between those dialects and the putative standard language. Further, it excludes these contrasts from the analysis of "standard language using" children's behavior. While it is notable that the children classed as speakers of the standard variety are so categorized owing to the relatively small amount of apparent dialectal variation they produce, the assessment that they categorically speak the standard variety has not been supported by a comprehensive theory predicting how standard and non-standard varieties are distinguished by the learner, and therefore has not reflected the differences in task complexity for learners with different linguistic experience.

Oetting et al. (2013) introduced the terms dialect-specific and dialect-universal to distinguish between features utilized by either or both of African American English (AAE) and Southern White English (SWE) speaking children. The authors promote a system-based approach to assessing child language development, stressing that non-mainstream dialects cannot be understood merely as variations on the mainstream defined standard language variety. Further work is needed to describe how child perception of social factors influences acquisition of both dialect specific and dialect universal features.

Similarly the question of how ranges of regional variation within ethnically defined dialects vary with respect to the regional variation of standard dialects remains largely unanswered. In order to accurately assess how dialectal variation emerges across populations of learners, it is wholly necessary to reevaluate how the construct of the "standard variety" is operationalized and why.

The question of how children with significant backgrounds in manifestly non-standard language varieties may nonetheless succeed at acquiring a standard variety cannot be adequately answered without a more precise assessment of which dialect(s) a child is acquiring, and how those dialects may be representationally related. Existing research on this topic largely focuses on school age children (e.g. Terry and Irving, 2010) but further work is required to define the role of dialect perception in early language acquistion.

In the next section I will discuss some evidence that measures of vocabulary size reflect learner beliefs about the relationship of their dialect to the standard variety. While larger vocabularies are associated with more favorable outcomes on a variety of measures, the assumption that vocabulary is itself an expression of linguistic capacity is not very well supported, but rather rests atop an age old bias against language varieties associated with socially less powerful groups.

6.4 Vocabulary as habitus

There is a large body of evidence linking early measures of lexical knowledge and long term outcomes. For example, increases in vocabulary size are correlated with greater evidence of segmental and phonological representations in the lexicon (Edwards, Beckman, and Munson, 2004; Janet F Werker and Curtin, 2005; Curtin and Werker, 2007). Experimental studies with monolingual English and Spanish learners demonstrate that early measures of proficiency in comprehending spoken language predict variability in later language outcomes (Anne Fernald, Perfors, and Marchman, 2006; Hurtado, Marchman, and Fernald, 2008). Larger vocabulary is also associated with improved performance on the Switch task (J. F. Werker et al., 2002). In this section I will outline how the affiliative model of language acquisition offers a clarifying framework in which to understand the correlation of desirable language outcomes with measures of vocabulary size.

The observation that greater vocabulary is associated with "better" outcomes has led many scientists and educators to make the generalization that larger vocabulary size is a characterization of the quality of the child's language, which in turn predicts the quality of their later language outcomes. This conclusion is also motivated by the finding that literacy outcomes are predicted by measures of phonological awareness. However, centering achievement in reading as a measure of language development again privileges the standard language variety and the social groups associated with it. Prioritizing literacy is, however uncontroversially so, an inherently political agenda (Rockhill, 1987). A purely scientific study of developmental speech processing must deconstruct popular assumptions about the comparative quality of language varieties and skills. By focusing on correcting the behavior of comparatively disadvantaged learners, proponents of deficit models aim to assist children in overcoming some qualitative poverty of stimulus. However, this assumption reveals linguistic bias - the assumption that vocabulary size is a *cause* of educational outcomes derives from a tempting logical fallacy. The commonplace characterization of larger vocabulary as an indication of 'better" language begs the question of why the various outcomes correlated with larger vocabulary are preferred.

Large scale intervention programs have been created to address educational inequity by targeting disadvantaged learners for additional instruction, or advising parents on how to better facilitate language learning (Hart and Risley, 2003; Marulis and Neuman, 2010) . The basis for such interventions is a theory which holds that effects of vocabulary size reflect a link between language exposure and language ability. Knowledge of a greater number of word forms is associated with more robustly generalized knowledge of lexical structure, potentially facilitating the ability to more quickly access the familiar patterns in other words. However, the assumption that populations with significantly different vocabulary sizes may be accurately categorized as speaking the same language is consistent with prescriptivist deficit models of language minority students. Such models postulate that differences which distinguish minority language students from non-standard language speakers are attributable to lower language proficiency, or semilingualism (MacSwan, 2000).

While linguistic differences are implicated in learning outcomes, they are far from the only factors known to impact the academic success of students. Baugh (2017) provides a detailed criticism of such efforts as a revival of the deficit model, especially with respect to African American English. The linguistic structures of by non-standard dialects are ill captured by a study which focuses on describing lexical abilities while excluding the study of phonetic, phonological, morphological, syntactic or semantic representations. Existing research programs risk under-specifying the distinction between language populations with different vocabularies. Contrastively, by asserting that distinct speaker affiliations demand distinct sociolinguistic competences, we can rigorously define linguistic standards and interpret measures of vocabulary with respect to them.

Rather than assuming that assessments of vocabulary directly measure language knowledge, and that language knowledge in turn impacts later linguistic and socioeconomic outcomes, we must be able to distinguish this scenario from one where the assessment reflects a common non-linguistic factor which independently predicts both early vocabulary scores and later outcomes. Of course, the premise of the interventionist approach to addressing minority language learners relies on the very observation that differences in learners' socioeconomic affiliation reliably correlate with differences in language measurements within and across different ethnic groups (Dollaghan et al., 1999; Pan, Rowe, Singer, and Snow, 2005; Hoff, 2006). Further, we must be sensitive to the potential for linguistic bias. I propose that the affiliative model presents an elegant solution to both problems by explicitly defining language as existing within a system of intrinsically inequitable social evaluation.

Effectively, this theory assumes that the vocabulary items which distinguish two learners' lexicons are necessarily not categorized as belonging to the same socially grounded linguistic competence. Supposing that both standardized and less prestigious speech forms are encoded with identical referential content, and a learner knows that both the forms "dag" and "dog" may refer to the same exact set of animals. Is it reasonable to assume they are also encoded as a single lexical item? Furthermore, what effect should we expect of experience with different dialects on this encoding? A learner who recognizes both forms as referring to the same object may have a myriad of distinct social beliefs about the kinds of people who use each speech form. The assumption that referential content alone is sufficient to define lexemes does not admit the possibility that social inferences impact lexical access.

The findings supporting a role of lexical knowledge in listener perceptual adaptation cannot be ignored, but they also must not be overstated. Evidence correlating vocabulary size with other language abilities is not necessarily evidence of a directional causative link between them. The affiliative model predicts that linguistic learning will also be subject to biases regarding the value of informants who appear to be potentially non-linguistically affiliated with different groups. It is necessary to account for the potential impact of cultural differences on vocabulary size, and this model challenges the assumption that interventions seeking to aid typically developing children with linguistic disadvantages ought to target the children's learning rather than the attitudes they encounter towards their language in the learning environment.

Rather than assuming knowledge of more word forms directly causes improved outcomes, we should interpret the utility of vocabulary growth in a social context. Positing a role of social interaction in vocabulary growth is not controversial (Hoff, 2003; Iverson and Goldin-Meadow, 2005; D. R. Powell, Son, File, and Froiland, 2012). Therefore, variation in vocabulary size can be understood as a measure of social stratification: possession of additional word forms correlates with the ability to communicate with more diverse listeners. Accordingly, large vocabularies may be associated with class mobility. In this way, measures of linguistic ability are predicted to behave as measures of general intelligence do, as markers of social class membership. I theorize it is not the lexicon itself but the learner's proclivity for identifying superior informants, and the associated linguistic habitus which drives lexical learning.

6.5 Conclusions

In this dissertation I have shown evidence that preverbal infants use inferences about relative knowledgeability to guide their acquisition of novel words. Further, this model can also account for effects of beliefs about group membership on infant responses to true and false labeling actions.

Further, I have shown that uniting accounts of selective trust with language learning has the potential to deepen our understanding of many areas of linguistic study. A research program in early sociophonetic learning has the potential to increase understanding of variation in language outcomes owing to differences in cultural background, identity, and disordered language skills. In applied linguistics, it may assist in understanding the etiology of academic achievement gaps, or functional differences between typically developing and developmentally disabled language users.

Data from knowledgeable speakers accurately identified as such aids in the prediction of both what referential content *other* knowledgeable speakers provide *and* what referential content they will not. Conversely, data from unknowledgeable speakers is not helpful in predicting the referential content of labels from either knowledgeable or other unknowledgeable speakers. In effect, knowledgeable speakers are not only expected to produce reliable speech, they are also expected to produce speech which can be reliably recognized as reliable, inducing in the listener not only a greater degree of confidence in their interpretation of the speech, but also a greater degree of confidence in the metalinguistic framework which produces that interpretation.

I submit that it is not the acquisition of lexical items, but a metalinguistic framework for reconciling competing hypotheses about the identity of lexical items that is the object of linguistic learning. Rather than relying on judgements of certainty about referential content, the learner's attention is distributed according to beliefs about the source and their accuracy. Therefore, we may describe language learners as relying on differences in the relative perception of uncertainty to select linguistic informants. Differences in the breadth or specificity of listener preferences for speakers in turn predict differences in task complexity, both within groups and at the level of individual differences.

This model allows us to define the acquisition task for mono-dialectal and multi-dialectal speakers of a common language as distinct. As learners exposed to two dialects develop preferences among them, we may accordingly predict differences in their linguistic representations depending on which of the two individual varieties they are acquiring, or whether they are acquiring both. Supposing multi-dialectal children must disambiguate utterances which plausibly could be produced by speakers of multiple dialects, the acquisition of additional dialects is expected to correlate with an increase in cognitive load.

In effect, the task of acquiring multiple linguistic codes is expected to increase in difficulty the more similarity the codes bear to one another. Speakers of multiple dialects are consequently expected to be more sensitive to variation in non-referential cues between speakers.

While disadvantaged students may benefit from interventions designed to increase their skills in standard varieties of speech, as language scientists we must not lose sight of the fact that linguistic behavior is socially situated, even for very young learners. To the extent that standard language skills are required for academic success, it is important to recognize that institutions are often fundamentally hostile to non-standard dialect-using children's native language varieties. While it is not enough to declare divergent developmental trajectories as equally valid, it is a disservice to conceptualize children's progress as falling short of standardized expectations without a principled and defensible definition of how these standards are operationalized. Equity in education cannot be achieved without addressing systemic biases which discriminate against non-standard language varieties and associated speaker identities and which are pervasive even in the scientific literature on this subject.

Chapter 7

Appendices

7.1 Simulation: Rost and McMurray (2009)

7.1.1 Joint inference of (C, K) for a single data point

The posterior distribution over C and K is different for different kinds of observations. Tables 7.2 - 7.4 give the unnormalized distribution after an observation of a single data point. Table 7.5 gives the normalized distribution for multiple data points.

Table 7.1: Definition of variables

C_x	category with index x
D_z	unambiguous data indicating an intention to articulate category C_z
w	knowledgeability
m	the number of observations
m_0	the number of observations of ambiguous data D_0
m_x	the number of observations of unambiguous data D_x indicating an intention to represent category C_x
m_b	the number of observations of unambiguous data D_b indicating an intention to represent some category $C_{b \neq x}$

The following table gives the unnormalized probability of observing of evidence for category x and knowledgeability w given that the category is C_1. Note that observing unambiguous evidence for a category other than the true category given a knowledgeable speaker has a probability of zero.

Table 7.2: C, K given unambiguous observation suggesting C_1

x	w	$P(C_x, K_w \mid D_1) = P(C_x, K_w) \sum_y P(D_1 \mid C_x, K_w, R_y) P(R_y)$
1	1	$c_1 k[(p_1)(1-r) + (q_1)r]$
2	1	0
3	1	0
1	0	$\frac{1}{n} c_1 (1-k)[p_1(1-r) + q_1 r]$
2	0	$\frac{1}{n} c_1 (1-k)[p_1(1-r) + q_1 r]$
3	0	$\frac{1}{n} c_1 (1-k)[p_1(1-r) + q_1 r]$

The following table gives the unnormalized probability of observing of evidence for category x and knowledgeability w given that the category is C_2.

Table 7.3: C, K given unambiguous observation suggesting C_2

x	w	$P(C_x, K_w) \sum_y P(D_2 \mid C_x, K_w, R_y) P(R_y)$
1	1	0
2	1	$c_2 k[(p_2)(1-r) + (q_2)r]$
3	1	0
1	0	$\frac{1}{n} c_2 (1-k)[p_2(1-r) + q_2 r]$
2	0	$\frac{1}{n} c_2 (1-k)[p_2(1-r) + q_2 r]$
3	0	$\frac{1}{n} c_2 (1-k)[p_2(1-r) + q_2 r]$

The following table gives the unnormalized probability of observing of evidence for category x and knowledgeability w given that the category is C_3.

Table 7.4: C, K given unambiguous observation suggesting C_3

x	w	$P(C_x, K_w) \sum_y P(D_3 \mid C_x, K_w, R_y) P(R_y)$
1	1	0
2	1	0
3	1	$c_3 k[(p_3)(1-r) + (q_3)r]$
1	0	$\frac{1}{n} c_3 (1-k)[p_3(1-r) + q_3 r]$
2	0	$\frac{1}{n} c_3 (1-k)[p_3(1-r) + q_3 r]$
3	0	$\frac{1}{n} c_3 (1-k)[p_3(1-r) + q_3 r]$

Next, we generalize the case of a single observation to a case of m distinct observations. The following table gives the normalized probability of observing of evidence for category x and knowledgeability w given that the category is C_x. The m observations may either be completely ambiguous ($m = m_0$), a mix of unambiguous and ambiguous data supporting the correct conclusion ($m = m_x + m_0$), or a mix of ambiguous and unambiguous data supporting an incorrect conclusion ($m = \sum_b m_b + m_0$).

Table 7.5 gives the generalized normalized joint distribution of the category and the knowledgeability of a single speaker supplying multiple data points. In the next section we will derive this formula for the special case of two observations from a single speaker.

Table 7.5: C, K given multiple observations

$condition$	K_w	$P(C_x, K_w \vert \vec{D}_z)$
$m_0 = m$	1	$\frac{c_x k[(1-r)(1-p_x)^{m_0}+r(1-q_x)^{m_0}]}{\sum_{x'} c_{x'} k[(1-r)(1-p_x)^{m_0}+r(1-q_x)^{m_0}]+\frac{(1-k)}{n}[\sum_{x''}(1-r)(1-p_{x''})^{m_0}+r(1-q_{x''})^{m_0}]}$
	0	$\frac{\frac{(1-k)}{n}[\sum_{x'}(1-r)(1-p_{x'})^{m_0}+r(1-q_{x'})^{m_0}]}{k[(1-r)(1-p_x)^{m_0}+r(1-q_x)^{m_0}]+\frac{(1-k)}{n}[\sum_{x'}(1-r)(1-p_{x'})^{m_0}+r(1-q_{x'})^{m_0}]}$
$m = m_x + m_0$	1	$\frac{nk}{[1+(n-1)k]}$
	0	$\frac{(1-k)}{[1+(n-1)k]}$
$m = \sum_b m_b + m_0$	1	1
	0	0

7.1.2 Joint inference of (C, K) for multiple data points from a single speaker

Table 7.6 gives the unnormalized distribution after a joint observation of two distinct data points. The two observations may either be both ambiguous, both unambiguously correct, both unambiguously incorrect and matching, or a mix of unambiguous and ambiguous data which supports either the correct or an incorrect conclusion. Because in this simulation each speaker samples from only one category, it is not possible to observe two unambiguous but mismatching tokens from a single speaker. This is reflected in the zero probability given in the last four lines of the table.

Table 7.6: unnormalized $P(C_x, K_w|\vec{D}_z)$

D_1	D_2	K_w	$P(C_x, K_{\vec{w}}) \sum_y P(\vec{D}_z|C_x, K_w)P(R_y)$
0	0	1	$c_x k[(1-r)(1-p_x)^2 + r(1-q_x)^2]$
0	0	0	$c_x(1-k)\frac{1}{n}[\sum_{x'}(1-r)(1-p_{x'})^2 + r(1-q_{x'})^2]$
$a = x$	$a = x$	1	$c_a k[(1-r)p_a^2 + rq_a^2]$
$a = x$	$a = x$	0	$c_a \frac{(1-k)}{n}[(1-r)p_a^2 + rq_a^2]$
$b \neq x$	$b \neq x$	1	0
$b \neq x$	$b \neq x$	0	$c_x \frac{(1-k)}{n}[(1-r)p_b^2 + rq_b^2]$
0	$a = x$	1	$c_a k[(1-r)(1-p_a)p_a + r(1-q_a)q_a]$
0	$a = x$	0	$c_a \frac{(1-k)}{n}[(1-r)(1-p_a)p_a + r(1-q_a)q_a]$
0	$b \neq x$	1	0
0	$b \neq x$	0	$c_x \frac{(1-k)}{n}[(1-r)(1-p_b)p_b + r(1-q_b)q_b]$
$a = x$	$b \neq x$	1	0
$a = x$	$b \neq x$	0	0
$b \neq x$	$c \neq b$	1	0
$b \neq x$	$c \neq b$	0	0

The foregoing distribution may be normalized by dividing each entry by a normalizing constant, determined. Table 7.7 gives these constants for each of five different conditions. The listener may either make two ambiguous observations, two correct or two incorrect and matching unambiguous observations, or one ambiguous observation paired with either a correct or an incorrect unambiguous observation.

Table 7.7: normalizing constants

D_{z_1}	D_{z_2}	$\sum_{x,w} P(C_x, K_w) \sum_{y'} P(\vec{D}_z \mid R_y, C_x, K_w) P(R_y)$
0	0	$\sum_x c_x[k[(1-r)(1-p_x)^2 + r(1-q_x)^2]+$ $(1-k)\frac{1}{n}[\sum_{x'}(1-r)(1-p_{x'})^2 + r(1-q_{x'})^2]]$
$a = x$	$a = x$	$[k + \frac{(1-k)}{n}] \sum_x c_x[(1-r)p_x^2 + rq_x^2]$
$b \neq x$	$b \neq x$	$\frac{(1-k)}{n} \sum_{x,b} c_x[(1-r)p_b^2 + rq_b^2]$
0	$a = x$	$[k + \frac{(1-k)}{n}] \sum_x c_x[(1-r)(1-p_x)p_x + r(1-q_b)q_x]$
0	$b \neq x$	$\sum_x c_x[k[(1-r)(1-p_x)p_x + r(1-q_b)q_x]+$ $\frac{(1-k)}{n} \sum_b[(1-r)(1-p_b)p_b + r(1-q_b)q_b]]$

Each line in table 7.6 divided by the corresponding constant given in table 7.7 provides a formula for the normalized probability distribution for category and knowledgeability given a pair of observations and a hypothesis about the true category (x).

Generalizing from the case of two observations, the unnormalized joint distribution of C and K for m observations is given in table 7.8. The total number of observations m is the sum of the number of ambiguous observations m_0, and the number of unambiguous observations made in support of either the correct hypothesis m_x or each incorrect hypothesis $\sum_b m_b$.

Table 7.8: unnormalized $P(C_x, K_w|\vec{D_z})$

condition	K_w	$P(C_x, K_{\vec{w}}) \sum_y (\vec{D_z}|C_x, K_w)P(R_y)$
$m = m_0$	1	$c_x k[(1-r)(1-p_x)^{m_0} + r(1-q_x)^{m_0}]$
$m = m_0$	0	$c_x \frac{(1-k)}{n}[\sum_{x'}(1-r)(1-p_{x'})^{m_0} + r(1-q_{x'})^{m_0}]$
$m = m_x$	1	$c_x k[(1-r)p_x^{m_x} + rq_x^{m_x}]$
$m = m_x$	0	$c_x \frac{(1-k)}{n}[(1-r)p_x^{m_x} + rq_x^{m_x}]$
$m = m_b$	1	0
$m = m_b$	0	$c_x \frac{(1-k)}{n}[(1-r)p_b^{m_b} + rq_b^{m_b}]$
$m_0, m_x > 0$	1	$c_x k[(1-r)(1-p_x)^{m_0} p_x^{m_x} + r(1-q_x)^{m_0} q_x^{m_a}]$
$m_0, m_x > 0$	0	$c_x \frac{(1-k)}{n}[(1-r)(1-p_x)^{m_0} p_x^{m_x} + r(1-q_x)^{m_0} q_x^{m_x}]$
$m_0, m_b > 0$	1	0
$m_0, m_b > 0$	0	$c_x \frac{(1-k)}{n}[(1-r)(1-p_b)^{m_0} p_b^{m_b} + r(1-q_b)^{m_0} q_b^{m_b}]$
$m_0, m_x, m_b > 0$	1	0
$m_0, m_x, m_b > 0$	0	0

7.1.3 Approximation of the Joint Inference of $(C, K|D)$

As the number of speakers grows beyond one, the number of possible combinations of speaker attributes grows exponentially. In order to estimate the joint distribution of C and K, we use Gibb's sampling, alternately estimating the conditional distributions of each variable given the observations $\vec{D}$.

Litmus Test: Inference of $(K|C,D)$

The Litmus Test is the name I have given to the estimation of the speaker's knowledgeability given an observation and a hypothesis about the category, as the outcome indicates whether the speaker should be trusted.

The following table gives the unnormalized probability of the speaker being knowledgeable supposing a belief that the category C has value x, and the listener has observed two labeling instances.

Table 7.9: unnormalized $P(K_w|C_X, D_{z_1}, D_{z_2})$

$[D_{z_1}, D_{z_2}]$	w=1	w=0
$0,0$	$k[(1-r)(1-p_x)^2 + r(1-q_x)^2]$	$\frac{(1-k)}{n}\sum_{x'}[(1-r)(1-p_{x'})^2 + r(1-q_{x'})^2]$
a,a	$k[(1-r)p_a^2 + rq_a^2]$	$\frac{(1-k)}{n}[(1-r)p_a^2 + rq_a^2]$
b,b	0	$\frac{(1-k)}{n}[(1-r)p_b^2 + rq_b^2]$
$0,a$	$k[(1-r)(1-p_a)p_a + r(1-q_a)q_a]$	$\frac{(1-k)}{n}[(1-r)(1-p_a)p_a + r(1-q_a)q_a]$
$0,b$	0	$\frac{(1-k)}{n}[(1-r)(1-p_b)p_b + r(1-q_b)q_b]$
a,b	0	0
b,c	0	0

Generalizing to m data points, Table 7.10 gives the unnormalized probability of the speaker being knowledgeable supposing a belief that the category C has value x, and the listener has observed a vector of observations $\vec{D_z}$. The unnormalized probability that the speaker is not knowledgeable is given in the following table, Table 7.11.

Table 7.10: unnormalized $P(K_1|C_X, \vec{D_z})$

condition	w=1
$m = m_0$	$k \sum_{x'}[(1-r)(1-p_{x'})^{m_0} + r(1-q_{x'})^{m_0}]$
$m = m_x$	$k[(1-r)(p_x)^{m_x} + r(q_x)^{m_x}]$
$m = m_b$	0
$m = m_0 + m_x$	$k[(1-r)(1-p_X)^{m_0} p_x^{m_x} + r(1-q_x)^{m_0} q_x^{m_x}]$
$m = m_0 + m_b$	0

Table 7.11: unnormalized $P(K_0|C_X, \vec{D_z})$

condition	w=0
$m = m_0$	$\frac{(1-k)}{n} \sum_{x'}[(1-r)(1-p_{x'})^{m_0} + r(1-q_{x'})^{m_0}]$
$m = m_x$	$\frac{(1-k)}{n}[(1-r)(p_x)^{m_x} + r(q_x)^{m_x}]$
$m = m_b$	$\frac{(1-k)}{n}[(1-r)(p_x)^{m_b} + r(q_x)^{m_b}]$
$m = m_0 + m_x$	$\frac{(1-k)}{n}[(1-r)(1-p_x)^{m_0} p_x^{m_x} + r(1-q_x)^{m_0} p_x^{m_x}]$
$m = m_0 + m_b$	$\frac{(1-k)}{n}[(1-r)(1-p_b)^{m_0} p_b^{m_b} + r(1-q_b)^{m_0} p_b^{m_b}]$

Trust Fall: Inference of $(C|K, D)$

I have termed the estimation of the category given an observation of a label and a hypothesis about the speaker's knowledgeability, as the outcome is conditioned on a belief about whether the speaker should be trusted.

Table 7.12 gives the unnormalized probability of a hypothesis about the category being correct, supposing that the speaker is knowledgeable. The unnormalized probability that a hypothesis about the category is correct supposing that the speaker is unknowledgeable is given in Table 7.13.

Table 7.12: $P(C_X|K_1, \vec{D}_z)$

condition	unnormalized	normalized
$m = m_0$	$c_x[(1-r)(1-p_x)^{m_0} + r(1-q_x)^{m_0}]$	$\frac{c_x[(1-r)(1-p_x)^{m_0}+r(1-q_x)^{m_0}]}{\sum_{x'} c_{x'}[(1-r)(1-p_{x'})^{m_0}+r(1-q_{x'})^{m_0}]}$
$m = m_{a=x}$	$c_x[(1-r)p_x^{m_x} + rq_x^{m_x}]$	1
$m = m_0 + m_{a=x}$	$c_x[(1-r)(1-p_x)^{m_0}p_x^{m_x} + r(1-q_x)^{m_0}q_x^{m_x}]$	1
$m = m_{b\neq x}$	0	0
$m = m_0 + m_{b\neq x}$	0	0

Table 7.13: $P(C_X|K_0, \vec{D}_z)$

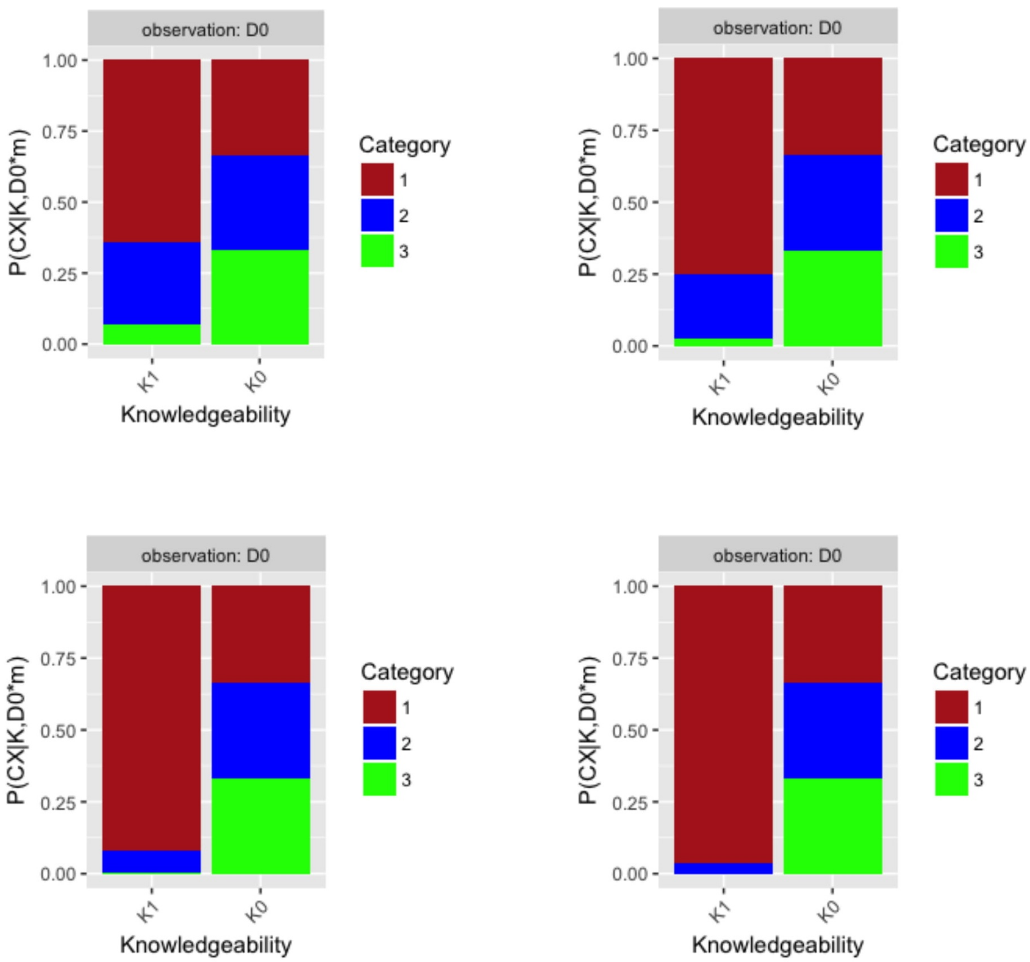
condition	unnormalized	normalized
$m = m_0$	$c_x\frac{1}{n}\sum_{x'}[(1-r)(1-p_{x'})^{m_0} + r(1-q_{x'})^{m_0}$	c_x
$m = m_{a=x}$	$c_x\frac{1}{n}[(1-r)p_x^{m_x} + rq_x^{m_x}]$	c_x
$m = m_{b\neq x}$	$c_x\frac{1}{n}[(1-r)p_b^{m_b} + rq_b^{m_b}]$	c_x
$m = m_0 + m_{a=x}$	$c_x\frac{1}{n}[(1-r)(1-p_x)^{m_0}p_x^{m_x} + r(1-q_x)^{m_0}q_x^{m_x}]$	c_x
$m = m_0 + m_{b\neq x}$	$c_x\frac{1}{n}[(1-r)(1-p_b)^{m_0}p_b^{m_b} + r(1-q_b)^{m_0}q_b^{m_b}]$	c_x

Trust Falls with Ambiguous Data

The model supposes that the listener distinguishes between ambiguous and unambiguous data, and therefore we predict that perception of a single disagreement between speakers is sufficient to result in the belief that one of the utterances is definitively incorrect. The inference that at least one of the observed labels is not correct additionally licenses an inference about the relative quality of the speakers.

Conversely, when perceiving strictly ambiguous data, growing numbers of observations are predicted to differently impact the posterior probability of knowledgeability depending on the prior distribution of the hypotheses. Ambiguous data does not deterministically signal a contrast between knowledgeable and unknowledgeable informants, and inferences about which informants are knowledgeable, and therefore independently, do not help the listener predict a speaker's status. However, additional ambiguous observations provide increasing evidence for the category which is most often realized ambiguously. The graphs in Figure 7.1 show the posterior probability of the category after increasing observations of ambiguous data, showing an increasing belief that initially favored hypothesis is correct.

Figure 7.1: $P(C_X|K_w, \vec{D_A})$
"Trust Fall given 2,3,6 and 8 ambiguous observations"



7.1.4 Joint inference of (C, K) for multiple speakers

The joint posterior distribution of C and K depends on the set of observations. Tables 7.14 - 7.15 gives a formula for the unnormalized joint probability of observing speakers with knowledgeability is K_{w_1} and K_{w_2} after observing them provide two data points, D_{z_1} and D_{z_2}.

Table 7.14: $P(C_x, K_{\vec{w}}|\vec{D_z})$

D_{z_1}	D_{z_2}	K_{w_1}	K_{w_2}	$P(C_x, K_{\vec{w}}) \prod_z \sum_y P(D_z|C_x, K_w)P(R_y)$
0	0	1	1	$c_x k^2[(1-r)(1-p_x) + r(1-q_x)]^2$
0	0	1	0	$c_x[k\frac{(1-k)}{n}][(1-r)(1-p_x) + r(1-q_x)]*$ $\sum_{x'}[(1-r)(1-p_{x'}) + r(1-q_{x'})]$
0	0	0	1	$c_x[k\frac{(1-k)}{n}][(1-r)(1-p_x) + r(1-q_x)]*$ $\sum_{x'}[(1-r)(1-p_{x'}) + r(1-q_{x'})]$
0	0	0	0	$c_x[\frac{(1-k)}{n}\sum_{x'}(1-p_{x'})]^2$
x	x	1	1	$c_x[kp_x]^2$
x	x	1	0	$c_x k\frac{(1-k)}{n}p_x^2$
x	x	0	1	$c_x k\frac{(1-k)}{n}p_x^2$
x	x	0	0	$c_x[\frac{(1-k)}{n}p_x]^2$
b	b	1	1	0
b	b	1	0	0
b	b	0	1	0
b	b	0	0	$c_x[\frac{(1-k)}{n}p_b]^2$

Table 7.15: $P(C_x, K_{\vec{w}} | \vec{D_z})$

D_{z_1}	D_{z_2}	K_{w_1}	K_{w_2}	$P(C_x, K_{\vec{w}}) \prod_z \sum_y P(D_z \vert C_x, K_w) P(R_y)$
x	0	1	1	$c_x k^2 (1 - p_x) p_x$
x	0	1	0	$c_x k \sum_{x'} (1 - p_{x'}) [\frac{(1-k)}{n}] p_x$
x	0	0	1	$c_x [\frac{(1-k)}{n}][1 - p_x] k p_x$
x	0	0	0	$c_x [\frac{(1-k)}{n}]^2 \sum_{x'} (1 - p_{x'}) [p_x]$
b	0	1	1	0
b	0	1	0	0
b	0	0	1	$c_x k [1 - p_x][\frac{(1-k)}{n}] p_b$
b	0	0	0	$c_x [\frac{(1-k)}{n}]^2 \sum_{x'} [p_b]$
x	b	1	1	0
x	b	1	0	$c_x k \frac{(1-k)}{n} [p_x p_b]$
x	b	0	1	0
x	b	0	0	$c_x \frac{(1-k)}{n}^2 [p_x p_b]$
b	c	1	1	0
b	c	1	0	0
b	c	0	1	0
b	c	0	0	$c_x \frac{(1-k)}{n}^2 [p_b p_c]$

Table 7.16 gives the normalizing constants for each possible relationship between the observations. The two observations could be both ambiguous, both unambiguous and generated from matched intentions, a mix of ambiguous and unambiguous data, or both

data points may be unambiguous and generated from mismatched intentions.

Table 7.16: normalizing constants for $P(C_x, K_{\vec{w}}|\vec{D}_z)$

D_{z_1}	D_{z_2}	$\sum_x \sum_{\vec{w}} P(C_x, K_{\vec{w}}) \prod_z P(D_z\|C_x, K_w)$
0	0	$\sum_x c_x[k^2[(1-r)(1-p_x)+r(1-q_x)]^2+$ $2[k+\frac{(1-k)}{n}]\sum_{x'}[(1-r)(1-p_{x'})+r(1-q_{x'})]+$ $[\frac{(1-k)}{n}\sum_{x'}(1-p_{x'})]^2]$
a	a	$c_a p_a^2[k^2+2k[\frac{(1-k)}{n}]+[\frac{(1-k)}{n}]^2]+\sum_{b\neq a} c_b[\frac{(1-k)}{n}p_b]^2$
0	a	$c_a p_a[k^2(1-p_a)+[\frac{(1-k)}{n}k][\sum_{x'}(1-p_{x'})+(1-p_a)]+[\frac{(1-k)}{n}]^2[\sum_{x'}(1-p_{x'})]^2]+$ $\sum_b c_b p_a[\frac{(1-k)}{n}][[1-p_b]k+[\frac{(1-k)}{n}]\sum_{x'}(1-p_{x'})]$
a	b	$[c_a+c_b][p_a p_b][k\frac{(1-k)}{n}+\frac{(1-k)}{n}^2]+[p_a p_b][\frac{(1-k)}{n}]^2\sum_{x\neq a,b}[c_x]$
b	c	$[p_b, p_c][\frac{(1-k)}{n}]^2\sum_{x\neq b,c}[c_x]$

Figure 7.2 describes the probability of believing each hypothesis about the category supposing an a string of ambiguous observations followed by an unambiguous observation of data which supports each hypothesis. The posterior distribution of categories depends on whether the speaker is knowledgeable. Each unambiguous observation from a knowledgeable speaker, notated $DXK1$ on the x-axis of the graph, results in the listener believing that the indicated category is correct. The same observation in the context of an unknowledgeable speaker, notated $DXK0$, results in a uniform distribution over hypotheses.

Figure 7.2: $P(D|I, R_0)$ and $P(D|I, R_1)$
"Beliefs about category after observing ambiguous data"

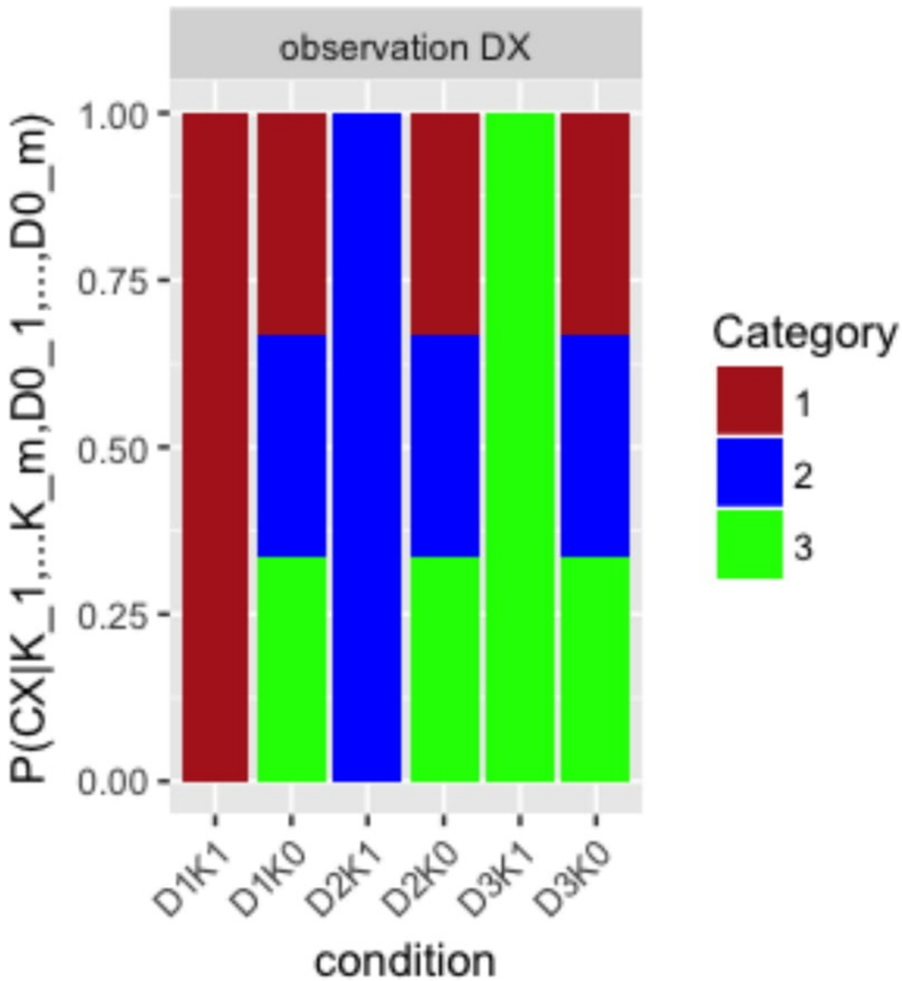
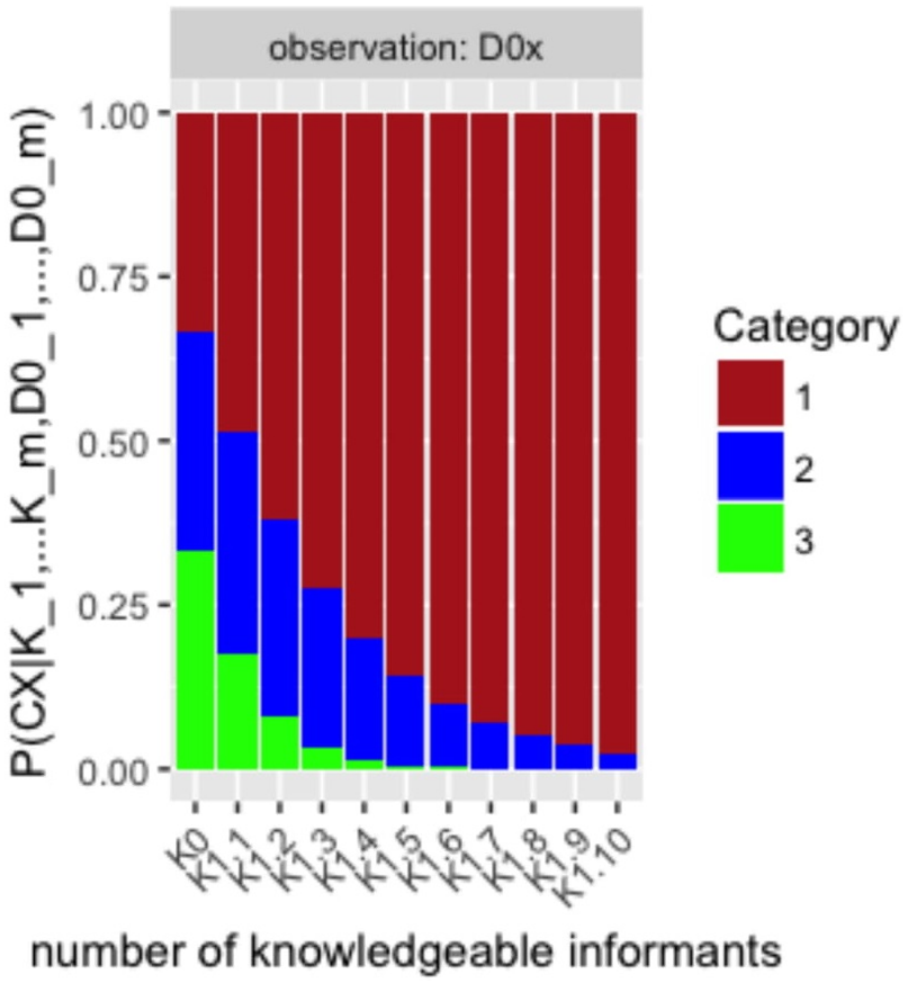


Figure 7.3 describes the probability of believing each hypothesis about the category supposing a string of ambiguous observations. The posterior distribution of categories depends on whether the speaker is knowledgeable. A string of ambiguous observations from a group of speakers who are all unknowledgeable, notated $K0$ on the x-axis of the graph, results in the listener believing that the posterior is identical to the uniform prior on categories. However, as increasing numbers of knowledgeable speakers are believed to be contributing, the posterior probability skews towards the category which is most often

realized ambiguously.

Figure 7.3: $P(D|I, R_0)$ and $P(D|I, R_1)$
"Beliefs about category after observing ambiguous data from multiple speakers"



Supposing all of the speakers are unknowledgeable, then the posterior on categories will be identical to the prior. Table 7.17 gives the probability of the category given two observations, one each from two different speakers, where *at least one* of the speakers is knowledgeable. Table 7.18 gives the normalizing constants for each of the three possible relationships between the observations - they may be both ambiguous, both unambiguous, or a combination.

Table 7.17: unnormalized $P(C_X|K_{w_1,w_2}, D_{z_1,z_2})$

$[D_{z_1}, D_{z_2}]$	$[K_{w_1,w_2}]$	unnormalized
$0, 0$	$1, 1$	$c_x[(1-r)(1-p_x)+r(1-q_x)]^2$
	$1, 0$	$c_x[(1-r)(1-p_x)+r(1-q_x)]$
	$0, 1$	$c_x[(1-r)(1-p_x)+r(1-q_x)]$
$a = X, a = X$	$1, 1$	$c_a[(1-r)(p_a)+r(q_a)]^2$
	$1, 0$	$c_a[(1-r)(p_a)+r(q_a)]$
	$0, 1$	$c_a[(1-r)(p_a)+r(q_a)]$
$0, a$	$1, 1$	$c_a[(1-r)(1-p_a)+r(1-q_a)][(1-r)(p_a)+r(q_a)]$
	$1, 0$	$c_a[(1-r)(1-p_a)+r(1-q_a)]$
	$0, 1$	$c_a[(1-r)(p_a)+r(q_a)]$

Table 7.18: normalizing constants for $P(C_X|K_{w_1,w_2}, D_{z_1,z_2})$

$[D_{z_1}, D_{z_2}]$	
$0, 0$	$\sum_x c_x[(1-r)(1-p_x)+r(1-q_x)]^2 + 2[(1-r)(1-p_x)+r(1-q_x)]$
a, a	$c_a[[(1-r)(p_a)+r(q_a)]^2 + 2[(1-r)(p_a)+r(q_a)]]$
$0, a$	$c_a[[(1-r)(1-p_a)+r(1-q_a)][(1-r)(p_a)+r(q_a)]+$ $[(1-r)(1-p_a)+r(1-q_a)]+[(1-r)(p_a)+r(q_a)]]$

To simplify, Table 7.18 combines the previous two tables to give the normalized probability of each category after two ambiguous, unambiguous or mixed observations. I use the notation T_a to reference the likelihood that the observation is ambiguously generated from an intention to represent category a, while H_a corresponds to the likelihood that the observation is unambiguously generated from this category.

Table 7.19: $P(C_X|K_{w_1,w_2}, D_{z_1,z_2})$

$[D_{z_1}, D_{z_2}]$	$[K_{w_1}, K_{w_2}]$	
$0, 0$	$1, 1$	$\frac{c_x[T_x]^2}{\sum_{x'} c_{x'}[T_{x'}]^2}$
$0, 0$	$1, 0$	$\frac{c_x[T_x]}{\sum_{x'} c_{x'}[T_{x'}]}$
$0, 0$	$0, 1$	$\frac{c_x[T_x]}{\sum_{x'} c_{x'}[T_{x'}]}$
a, a	$1, 1$	$\frac{c_a[H_a]^2}{c_a[H_a]^2}$
a, a	$1, 0$	$\frac{c_a[H_a]}{c_a[H_a]}$
a, a	$0, 1$	$\frac{c_a[H_a]}{c_a[H_a]}$
$0, a$	$1, 1$	$\frac{c_a[H_a][T_a]}{c_a[H_a][T_a]}$
$0, a$	$1, 0$	$\frac{c_a[T_a]}{c_a[T_a]}$
$0, a$	$0, 1$	$\frac{c_a[H_a]}{c_a[H_a]}$

7.2 Simulation: Koenig and Echols (2003)

Infants looked longest to human labelers who gave incorrect labels. We show that this looking behavior may correspond to higher uncertainty about the knowledgeability of human informants compared to audio speakers which directly follows from the assumption that human speakers are more often knowledgeable.

We have defined knowledgeability as dependent upon group membership, therefore we model listeners as having distinct beliefs about the patterns of speech informativity associated with observations from different groups. The Kullback-Leibler divergence describes the relative entropy of a distribution compared to a reference distribution. In this case, we will compare the average number of bits learned about the speaker's knowledgeability given a speech observation how much is learned about the speaker's knowledgeability without a speech observation. In effect, the KL divergence provides a comparative measure of how much the infant learns about the knowledgeability of different speaker types, assuming they use an encoding scheme which is optimized to explain variability in the speech data.

Assuming that human speakers are expected to be more knowledgeable than audio speakers, their speech will have a higher relative entropy than audio speakers, independent of the speech value. Assuming that human speakers who provide incorrect labels are less frequent than humans who provide correct labels, their speech will have a relatively higher entropy. From these two premises we can predict that infants attended longest to the human labelers who provided incorrect labels.

$$D_{KL}(P(K_w|D_d, G_g, C_x)||P(K_w|G_g, C_x)) = \sum_w P(K_w|D_d, G_g, C_x) \log \frac{P(K_w|D_d, G_g, C_x)}{P(K_w|G_g, C_x)} \tag{7.1}$$

Rewriting the recurring term of in equation 7.1 produces equation 7.2.

$$P(K_w|D_d, G_g, C_x) = \frac{P(D_d|K_w, G_g, C_x)P(K_w|G_g, C_x)}{\sum_{w'} P(D_d|K_{w'}, G_g, C_x)P(K_{w'}|G_g, C_x)} \tag{7.2}$$

Substituting equation 7.2 into the term from 7.1.

$$\frac{P(K_w|D_d, G_g, C_x)}{P(K_w|G_g, C_x)} = \frac{P(D_d|K_w, G_g, C_x)}{\sum_{w'} P(D_d|K_{w'}, G_g, C_x)P(K_{w'}|G_g, C_x)} \tag{7.3}$$

Substituting equations 7.2 and 7.3 into equation 7.1 yields the expression for KL divergence in 7.4.

$$\sum_w \frac{P(D_d|K_w, G_g, C_x)P(K_w|G_g, C_x)}{\sum_{w'} P(D_d|K_{w'}, G_g, C_x)P(K_{w'}|G_g, C_x)} \log \frac{P(D_d|K_w, G_g, C_x)}{\sum_{w'} P(D_d|K_{w'}, G_g, C_x)P(K_{w'}|G_g, C_x)} \tag{7.4}$$

Table 7.20: Components of KL Divergence between distribution of knowledgeability with and without observation of speech, in bits

D, C, G	w	$P(K_w\|G_g, C_x)$	$P(D_d\|K_w, G_g, C_x)$
D_x, x, audio speaker	0	0.85	$\frac{p_x}{3}$
	1	0.15	p_x
$D_{\neg x}$, x, audio speaker	0	0.85	$\frac{p_{\neg x}}{3}$
	1	0.15	0
D_x, x, adult	0	0.2	$\frac{p_x}{3}$
	1	0.8	p_x
$D_{\neg x}$, x, adult	0	0.2	$\frac{p_{\neg x}}{3}$
	1	0.8	0

Table 7.21: sum of possible outcomes

D, C, G	$\sum_{w'} P(D_d \mid K_{w'}, G_g, C_x) P(K_{w'} \mid G_g, C_x)$
D_x, x, audio speaker	$\frac{0.85p_x}{3} + 0.15p_x = 0.4333p_x$
$D_{\neg x}$, x, audio speaker	$\frac{0.85p_{\neg x}}{3} + 0 * 0.15 = 0.2833p_{\neg x}$
D_x, x, adult	$\frac{0.2p_x}{3} + 0.8p_x = 0.8667p_x$
$D_{\neg x}$, x, adult	$\frac{0.2p_{\neg x}}{3} + 0 * 0.8 = 0.0667p_{\neg x}$

Table 7.22: Formula for KL divergence of K with D compared to without for each type of observation

D, C, G	w	
D_x, x, audio speaker	0	$\frac{\frac{0.85p_x}{3}}{0.4333p_x} \log \frac{\frac{p_x}{3}}{0.4333p_x}$
	1	$\frac{0.15p_x}{0.4333p_x} \log \frac{p_x}{0.4333p_x}$
$D_{\neg x}$, x, audio speaker	0	$\frac{\frac{0.85p_{\neg x}}{3}}{0.2833p_{\neg x}} \log \frac{\frac{p_{\neg x}}{3}}{0.2833p_{\neg x}}$
	1	$\frac{0}{0.2833p_{\neg x}} \log \frac{0}{0.2833p_{\neg x}}$
D_x, x, adult	0	$\frac{\frac{0.2p_x}{3}}{0.8667p_x} \log \frac{\frac{p_x}{3}}{0.8667p_x}$
	1	$\frac{0.8p_x}{0.8667p_x} \log \frac{p_x}{0.8667p_x}$
$D_{\neg x}$, x, adult	0	$\frac{\frac{0.2p_{\neg x}}{3}}{0.0667p_{\neg x}} \log \frac{\frac{p_{\neg x}}{3}}{0.0667p_{\neg x}}$
	1	$\frac{0}{0.0667p_{\neg x}} \log \frac{0}{0.0667p_{\neg x}}$

Table 7.23: KL divergence of K with D compared to without for each type of observation, in bits

D, C, G	w	
D_x, x, audio speaker	0	-0.2474
	1	0.4176
	$\sum_w$	0.1701
$D_{\neg x}$, x, audio speaker	0	$\log \frac{1}{0.85}$
	1	0
	$\sum_w$	0.2345
D_x, x, adult	0	-0.1060
	1	0.1906
	$\sum_w$	0.0846
$D_{\neg x}$, x, adult	0	$\log \frac{1}{0.2}$
	1	0
	$\sum_w$	2.3219

7.3 Inferring informant quality based on non-linguistic affiliative cues

The probability that a person is knowledgeable given that we have observed their shirt to be red is the sum of the probability that they are a knowledgeable employee with a red shirt, or a knowledgeable customer with a red shirt, as given in equation 7.5.

$$P(K=1|F_1) = \sum_G P(K|G)P(G|F_1) \tag{7.5}$$

Using Bayes theorem, we rewrite the posterior probability on the latent employment feature G when F_1 is known in equation 7.6 .

$$P(L|F_1) = \frac{P(F|L)P(L)}{\sum_{G'} P(F_1|G')P(G')} \tag{7.6}$$

Combining equations 7.5 and 7.6, the posterior probability on knowledgeability after observing whether the informant has a red shirt is given in equation 7.7 below.

$$P(K=1|F) = \sum_G \frac{P(K_1|G)P(F|G)P(G)}{\sum_{G'} P(F_1|G')P(G')} \tag{7.7}$$

The posterior probability on unknowledgeability after observing whether the informant has a red shirt is a distinct distribution over groups, given in equation 7.8 below.

$$P(K=0|F) = \sum_G \frac{P(K_0|G)P(F|G)P(G)}{\sum_{G'} P(F_1|G')P(G')} \tag{7.8}$$

Importantly, the conditional probability of knowledgeability depends on beliefs about all groups, not only the group currently being observed. The expectation that a speaker is knowledgeable is therefore only conditionally independent of the category, given their membership in some group G.

References

Akhtar, N., & Gernsbacher, M. A. (2008). On privileging the role of gaze in infant social cognition. *Child development perspectives*, *2*(2), 59–65.

Apfelbaum, K. S., & McMurray, B. [B.]. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*(6), 1105–1138.

Balaban, M. T., & Waxman, S. R. [Sandra R]. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of experimental child psychology*, *64*(1), 3–26.

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.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child development*, *68*(1), 48–57.

Baugh, J. (1988). Language and race: Some implications for linguistic science. *Linguistics: the Cambridge survey*, *4*, 64–74.

Baugh, J. (2017). Meaning-less differences: Exposing fallacies and flaws in "the word gap" hypothesis that conceal a dangerous "language trap" for low-income american families and their children. *International Multilingual Research Journal*, *11*(1), 39–51.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental science*, *8*(6), 492–499.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, *30*(3), 359–375.

Bradbury, B., Corak, M., Waldfogel, J., & Washbrook, E. (2015). *Too many children left behind: The us achievement gap in comparative perspective*. Russell Sage Foundation.

Brooker, I., & Poulin Dubois, D. (2013). Is a bird an apple? the effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy*, *18*, E46–E68.

Brooks-Gunn, J., Rouse, C., & McLanahan, S. (2007). Racial and ethnic gaps in school readiness. *School readiness and the transition to kindergarten in the era of accountability.* 283–306.

Butler, J., Floccia, C., Goslin, J., & Panneton, R. (2011). Infants' discrimination of familiar and unfamiliar accents in speech. *Infancy*, *16*(4), 392–417.

Buttelmann, D. [D.], Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in group over out group members in 14?month?old infants. *Child Development*, *84*, 422–428.

Buttelmann, D. [David], & Böhm, R. (2014). The ontogeny of the motivation that underlies in-group bias. *Psychological Science*, *25*(4), 921–927.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior and development*, *21*(2), 315–330.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335–344.

Chonchaiya, W., & Pruksananonda, C. (2008). Television viewing associates with delayed language development. *Acta Paediatrica*, *97*(7), 977–982.

Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental science*, *11*(5), 761–770.

Cirelli, L. K., Wan, S. J., & Trainor, L. J. (2016). Social effects of movement synchrony: Increased infant helpfulness only transfers to affiliates of synchronously moving partners. *Infancy*, *21*(6), 807–821.

Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of american english regional dialects. *journal of phonetics*, *32*(1), 111–140.

Corriveau, K., & Harris, P. L. [P. L.]. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental science*, *12*(3), 426–437.

Corriveau, K. H. [Kathleen H], Kurkul, K., & Arunachalam, S. (2016). Preschoolers' preference for syntactic complexity varies by socioeconomic status. *Child development*, *87*(5), 1529–1537.

Corriveau, M., K. Fusaro, & Harris, P. L. [P. L]. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological science*, *20*(3), 372–377.

Coulon, M., Guellai, B., & Streri, A. (2011). Recognition of unfamiliar talking faces at birth. *International Journal of Behavioral Development*, *35*(3), 282–287.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. processes of change in brain and cognitive development. *Attention and performance*, *21*(21), 249–274.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148–153.

Curtin, S., & Werker, J. F. [Janet F]. (2007). The perceptual foundations of phonological development. In *The oxford handbook of psycholinguistics*.

De Boer, B., & Kuhl, P. K. [Patricia K]. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, *4*(4), 129–134.

DeCasper, A. J., & Fifer, W. P. (2013). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*, 1174–1176.

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1432–1443.

Eaves, B. S., & Shafto, P. (2017). Parameterizing developmental changes in epistemic trust. *Psychonomic Bulletin and Review*, *24*, 277–306.

Eckert, P. (2008). Variation and the indexical field 1. *Journal of sociolinguistics*, *12*(4), 453–476.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of speech, language, and hearing research*, *47*(2), 421–436.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.

Fecher, N., & Johnson, E. K. [Elizabeth K]. (2018). The native-language benefit for talker identification is robust in 7.5-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Fennell, C. T. (2012). Object familiarity enhances infants use of phonetic detail in novel words. *Infancy*, *17*(3), 339–353.

Fennell, C. T., & Waxman, S. R. [S. R.]. (2006). Infants of 14 months use phonetic detail in novel words embedded in naming phrases. In *Proceedings of the 30th annual boston university conference on language development* (Vol. 1, pp. 178–189). Somerville, MA: Cascadilla Press.

Fennell, C. T., & Waxman, S. R. [S. R.]. (2010). What paradox? referential cues allow for infant use of phonetic detail in word learning. *Child development*, *81*(5), 1376–1383.

Fennell, C. T., Waxman, S. R., & Weisleder, A. (2007). With referential cues, infants successfully use phonetic detail in word learning. In *Proceedings of the 31st annual boston university conference on language development* (Vol. 1, p. 206). Somerville, MA: Cascadilla Press.

Fennell, C. T., & Werker, J. F. [J. F.]. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and speech*, *46*, 2–3.

Fennell, C. T., & Werker, J. F. [J. F.]. (2004). Infant attention to phonetic detail: Knowledge and familiarity effects. In *Proceedings of the 28th annual boston university conference on language development* (Vol. 1, pp. 165–176). Somerville, MA: Cascadilla Press.

Fernald, A. [A.], Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231.

Fernald, A. [Anne], 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.

Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*, *124*(1), 95–100.

Foulkes, P. [P.], Docherty, G., & Watt, D. (1999). Tracking the emergence of sociophonetic variation. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 1625–1628).

Foulkes, P. [Paul], & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of phonetics*, *34*(4), 409–438.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585.

Fried, L., & Holyoak, K. J. (1984). Probabilistic models of language processing and acquisition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*(2), 234–234.

Frumkin, L. (2007). Influences of accent and ethnic background on perceptions of eyewitness testimony. *Psychology, Crime & Law*, *13*(3), 317–331.

Fulkerson, A. L., & Haaf, R. A. [Robert A]. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9-and 15-month-olds' object categorization. *Infancy*, *4*(3), 349–369.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, *6*, 721–741.

Gergely, G., Bekkering, H., & Kiraly, I. (2002). Developmental psychology: Rational imitation in preverbal infants. *Nature*, *415*(6873), 755.

Gergely, G., Egyed, K., & Kiraly, I. (2007). On pedagogy. *Developmental Science*, *10*(1), 139–146.

Gliga, T., Elsabbagh, M., Hudry, K., Charman, T., Johnson, M. H., & Team, B. (2012). Gaze following, gaze reading, and word learning in children at risk for autism. *Child development*, *83*(3), 926–938.

Group", " G., Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., … Larsen-Freeman, D., et al. (2009). Language is a complex adaptive system: Position paper. *Language learning*, *59*, 1–26.

Guy, G. R. (1988). Language and social class. *Linguistics: The Cambridge Survey*, *4*, 37–63.

Gweon, H., Pelton, H., & Schulz, L. E. (2011). Adults and school-aged children accurately evaluate sins of omission in pedagogical contexts. In *In proceedings of the cognitive science society* (Vol. 33, *33*).

Halle, P. A., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, *17*(2), 119–129.

Harris, P. L. [P. L.], & Corriveau, K. H. [K. H.]. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1567), 1179–1187.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American educator*, *27*(1), 4–9.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The linguistic review*, *23*(3), 351–379.

Hehman, E., Volpert, H. I., & Simons, R. F. (2013). The n400 as an index of racial stereotype accessibility. *Social Cognitive and Affective Neuroscience*, *9*(4), 544–552.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child development*, *74*(5), 1368–1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental review*, *26*(1), 55–88.

Hosoda, M., Stone-Romero, E. F., & Walter, J. N. (2007). Listeners' cognitive and affective reactions to english speakers with standard american english and asian accents. *Perceptual and Motor Skills*, *104*(1), 307–326.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? links between maternal talk, processing speed and vocabulary size in spanish-learning children. *Developmental science*, *11*(6), F31–F39.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. (2010). Sources of variability in children's language growth. *Cogn Psychol*, *61*(4), 343–65.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, *16*(5), 367–371.

Jin, K. S., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. In *Proceedings of the national academy of sciences* (Vol. 114, *31*, pp. 8199–8204).

Johnson, E. K. [E. K.], Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*(5), 1002–1011.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants? detection of the sound patterns of words in fluent speech. *Cognitive psychology*, *29*(1), 1–23.

Kelly, D. J. [D. J.], C., Q. P., Slater, A. M., Lee, K., Gibson, A., Smith, M., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental science*, *8*(9).

Kelly, D. J. [David 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.

Kenward, B., Karlsson, M., & Persson, J. (2011). Over-imitation is better explained by norm learning than by distorted causal learning. In *Proceedings of the royal society of london b: Biological sciences* (Vol. 278, *1709*, pp. 1239–1246).

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2005). Children's selective trust in native-accented speakers. *Developmental science*, *14*(1), 106–111.

Kinzler, K. D., & Dautel, J. B. (2012). Children?s essentialist reasoning about language and race. *Developmental Science*, *15*(1), 131–138.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. (Vol. 104, *30*, pp. 12577–12580).

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.

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition*, *87*(3), 179–208.

Koenig, M. A., & Harris, P. L. [P. L.]. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277.

Kuhl, P. K. [P. K.]. (2007). Is speech learning 'gated' by the social brain?. *Developmental science*, *10*(1), 110–120.

Kuhl, P. K. [P. K.], Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science*, *9*(2), F13–F21.

Kuhl, P. K. [Patricia K]. (2011). Who's talking? *Science*, *333*(6042), 529–530.

Labov, W. (1991). The three dialects of english. *New ways of analyzing sound change*, *5*, 1–44.

Labov, W. (2006). *The social stratification of english in new york city*. Cambridge University Press.

Lane, H. L. (1992). *The mask of benevolence: Disabling the deaf community*. Knopf New York.

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing–evidence from event-related brain potentials. *Neuroscience letters*, *339*(3), 191–194.

Legare, C. H., & Harris, P. L. [Paul L]. (2016). The ontogeny of cultural learning. *Child Development*, *87*(3), 633–642.

Levy, G. D. [G. D.]. (2003). Perception of correlated attributes involving african?american and white females' faces by 10?month?old infants. *Infant and Child Development: An International Journal of Research and Practice*, *12*(2), 197–203.

Levy, G. D. [Gary D.], & Haaf, R. A. [Robert A.]. (1994). Detection of gender-related categories by 10-month-old infants. *Infant Behavior and Development*, *17*(2), 457–459.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, *54*(5), 358.

Liberman, Z., Kinzler, K. D., & Woodward, A. L. (2014). Friends or foes: Infants use shared evaluations to infer others? social relationships. *Journal of Experimental Psychology: General*, *143*(3), 966.

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). Preverbal infants infer third-party social relationships based on language. *Cognitive science*, *41*, 622–634.

Lisker, L., & Abramson, A. S. (1963). Crosslanguage study of voicing in initial stops. *The Journal of the Acoustical Society of America*, *35*(11), 1889–1890.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, *104*(50), 19751–19756.

MacSwan, J. (2000). The threshold hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities. *Hispanic Journal of Behavioral Sciences*, *22*(1), 3–45.

Mahajan, N., & Wynn, K. (2012). Origins of "us" versus "them": Prelinguistic infants prefer similar others. *Cognition*, *124*(2), 227–233.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of educational research*, *80*(3), 300–335.

Maurer, D., & Werker, J. F. [Janet F]. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178.

McGuigan, N., & Whiten, A. (2009). Emulation and overemulation in the social learning of causally opaque versus causally transparent tool use by 23-and 30-month-olds. *Journal of experimental child psychology*, *104*(4), 367–381.

McMurray, B. [Bob], Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental psychology*, *24*(4), 470.

Milroy, J., & Milroy, L. (2012). *Authority in language: Investigating standard english*. Routledge.

Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen-month-olds know what 'we' have shared in a special way. *Infancy*, *13*(1), 90–101.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500.

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. *Child Development*, *84*(6), 2064–2078.

Nardy, A., Chevrot, J.-P., & Barbu, S. (2013). The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics*, *51*(2), 255–284.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 756.

Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental psychology*, *47*(4), 1050.

Norris, D., & McQueen, J. M. (2008). Shortlist b: A bayesian model of continuous speech recognition. *Psychological review*, *115*(2), 357.

Oetting, J. B., Lee, R., & Porter, K. L. (2013). Evaluating the grammars of children who speak nonmainstream dialects of english. *Topics in language disorders*, *33*(2), 140.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*(3), 273–285.

Pajak, B., Bicknell, K., & Levy, R. (2013). A model of generalization in distributional learning of phonetic categories. In *Proceedings of the fourth annual workshop on cognitive modeling and computational linguistics (cmcl)* (pp. 11–20).

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child development*, *76*(4), 763–782.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, *18*(1), 79–85.

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language*, 384–402.

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*(6), 2988–2998.

Pinker, S. (1979). Formal models of language learning. *Cognition*, *7*(3), 217–283.

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring canadian english and canadian french: Native language, cross dialect, and cross language comparisons. *Infancy*, *17*(2), 198–232.

Poulin Dubois, D. [D.], Brooker, I., & Polonia, A. (2011). Infants prefer to imitate a reliable person. *Infant Behavior and Development*, *34*(2), 303–309.

Poulin-Dubois, D., & Chow, V. (2009). The effect of a looker's past reliability on infants' reasoning about beliefs. *Developmental Psychology*, *45*(6), 1576.

Powell, D. R., Son, S.-H., File, N., & Froiland, J. M. (2012). Changes in parent involvement across the transition from public school prekindergarten to first grade and children's academic outcomes. *The Elementary School Journal*, *113*(2), 276–300.

Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. In *Proceedings of the national academy of sciences* (E3965–E3972).

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, *97*(1), B13–B23.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on american english dialect identification. *Journal of language and social psychology*, *18*(1), 10–30.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*(9), 1109–1121.

Rasanen, O. (2012). Computational modeling of phonetic and lexical learning in early language acquisition: Existing models and future directions. *Speech Communication*, *54*(9), 975–997.

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*(3), 651.

Rockhill, K. (1987). Gender, language and the politics of literacy. *British Journal of Sociology of Education*, *8*(2), 153–167.

Rost, G. C., & McMurray, B. [B.]. (2009). Speaker variability augments phonological processing in early word learning. *Developmental science*, *12*(2), 339–349.

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental psychology*, *47*(1), 19.

Schmale, R., Cristia, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, *15*(6), 650–662.

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15*, 436–447.

Shafto, P., & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In *Proceedings of the 30th annual conference of the cognitive science society* (pp. 1632–1637).

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive psychology*, *71*, 55–89.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640.

Siyanova-Chanturia, A., Pesciarelli, F., & Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS One*, *7*(12), e48712.

Spokes, A. C., & Spelke, E. S. (2017). The cradle of social knowledge: Infants' reasoning about caregiving and affiliation. *Cognition*, *159*, 102–116.

Stager, C. L., & Werker, J. F. [J. F.]. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(5), 381.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological science*, *13*(5), 480–484.

Tenenbaum, J. B. [J. B.], & Griffiths, T. L. (2001). The rational basis of representativeness. In *In proceedings of the 23rd annual conference of the cognitive science society* (pp. 1036–41).

Terry, N. P., & Irving, M. A. (2010). Cultural and linguistic diversity: Issues in education. *Special education for all teachers*, *5*, 109–132.

Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants track the reliability of potential informants. *Psychological Science, 25*(9), 1730–1738.

Tuninetti, A., Chládková, K., Peter, V., Schiller, N. O., & Escudero, P. (2017). When speaker identity is unavoidable: Neural processing of speaker identity cues in natural speech. *Brain and language*, *174*, 42–49.

United States Census Bureau. (2015). Detailed languages spoken at home and ability to speak english for the population 5 years and over for united states: 2009-2013. http://www.census.gov.

Wagner, L., Clopper, C. G., & Pate, J. K. (2014). Children's perception of dialect variation. *Journal of child language*, *41*(5), 1062–1084.

Washington, J. A., & Craig, H. K. (1994). Dialectal forms during discourse of poor, urban, african american preschoolers. *Journal of Speech, Language, and Hearing Research*, *37*(4), 816–823.

Washington, J. A., Craig, H. K., & Kushmaul, A. J. (1998). Variable use of african american english across two language sampling contexts. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1115–1124.

Waxman, S. R. [Sandra R], & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive psychology*, *29*(3), 257–302.

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

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

Werker, J. F. [J. F.], & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*(1), 49–63.

Werker, J. F. [Janet F], & Curtin, S. (2005). Primir: A developmental framework of infant speech processing. *Language learning and development*, *1*(2), 197–234.

White, K. R., Crites Jr, S. L., Taylor, J. H., & Corral, G. (2009). Wait, what? assessing stereotype incongruities using the n400 erp component. *Social Cognitive and Affective Neuroscience*, *4*(2), 191–198.

Xiao, N. G., Wu, R., Quinn, P. C., Liu, S., Tummeltshammer, K. S., Kirkham, N. Z., & Lee, K. (2018). Infants rely more on gaze cues from own?race than other?race adults for learning under uncertainty. *Child development*, *89*(3), e229–e244.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223–250.

Xu, F., & Tenenbaum, J. B. [Joshua B]. (2007). Word learning as bayesian inference. *Psychological review*, *114*(2), 245.

Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. In *In proceedings of the national academy of sciences* (Vol. 105, *36*, pp. 13690–13695).

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen month old infants learn similar sounding words. *Developmental science*, *12*(3), 412–418.

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, *106*(4), 208–220.