EXPERIMENTAL INVESTIGATIONS OF SOCIOLINGUISTIC KNOWLEDGE

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Laura Staum

December, 2008

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

____

(Penelope Eckert) Principal Co-Advisor

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

____

(Arnold Zwicky) Principal Co-Advisor

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

____

(John R. Rickford)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

____

(Meghan Sumner)

Approved for the University Committee on Graduate Studies

____

Abstract

A series of six experiments addressed the question, what is the nature of sociolinguistic knowledge? Three sub-questions were investigated: First, how do listeners use facts about speech to inform their beliefs about speakers? Second, and conversely, how might listeners use facts about speakers to inform their perceptions of speech? Third, what is the nature of the representations that listeners form of the social conditioning of variation? Results of Experiments 1a and 1b demonstrate that listeners can infer characteristics of speakers from their use of individual sociolinguistic variables. Results of Experiments 2a and 2b show that listeners use social information about speakers to understand their speech. Results of Experiments 3a and 3b suggest that listeners form socially specified representations at some level during speech processing, but that their underlying phonological representations are the same for all speakers. Taken together, results of all six experiments show bidirectional influences of information in the speech stream on inferences about social characteristics of the speaker, and of social information on speech perception. I propose a Bayesian approach to integrating social information into a model of language comprehension, in which bidirectional influences between verbal and nonverbal factors emerge as a natural consequence of our cognitive capacities for learning and inference.

Acknowledgments

I have many people to thank for their contributions to this project. My committee has shaped this work in more ways than I can count. I owe a great deal of my outlook on the world of sociolinguistics to Penny Eckert, who has never been afraid of asking the questions nobody else is asking, and sees the forest and the trees at the same time. John Rickford has been a truly excellent role model of scholarship and teaching and research and life all rolled together, and his comments have always been both insightful and thorough. Many of the items and techniques appearing in the experiments are due to the creativity of Meghan Sumner, who never hesitated to get her hands dirty with me in the sandbox of experimental design. And special thanks go to Arnold Zwicky, for being a constant source of new ideas, cool examples, and positive energy, when all were sorely needed.

Many other people made intellectual contributions to this work, from discussing the original seed of the idea to reading over drafts of the final document. The Stanford Socio Dissertation Group (Katie Drager, Rebecca Greene, Lauren Hall-Lew, Stacy Lewis, and Rebecca Starr) were excellent critics, cheerleaders, and editors. Herb Clark, Liz Coppock, Evelina Fedorenko, Ted Gibson, and Dan Jurafsky, as well as audiences at SLUGs and SPLaT!, have all given helpful comments. Siobhan Greatorex-Voith, Ben de Jesus, and Vivian Chau provided enthusiastic and conscientious help preparing the stimuli and running subjects. On the financial side, I received research support from NSF Grant #BCS-0720054 to Meghan Sumner, and from a 2007 Mellon Dissertation Year Fellowship and a 2008 Stanford Graduate Research Opportunity Award, both to myself.

Finally, I owe the greatest debt of gratitude to Daniel Casasanto, whose intellectual, emotional, and practical contributions to this work cannot be overestimated. After this experience, I can confidently say that the only thing better than having a clear-thinking, experienced and talented experimentalist as your advisor or your collaborator is having one as your spouse.

Contents

List of Tables and Figures

Overview

This dissertation focuses on a question that is at the intersection of sociolinguistics and psycholinguistics: What do listeners know about sociolinguistic variation? In a series of six experiments, I addressed three sub-questions in service of this larger goal: How do listeners use facts about speech to make inferences about speakers? How might they also use facts about speakers to inform their perceptions of speech? And what is the nature of the representations that listeners form of the social conditioning of variation that underlie these operations?

Speech perception is difficult for several reasons, perhaps most notably because there is no one-to-one relationship between sounds and phonemes. The same sound can correspond to different categories in different contexts when spoken by different speakers. Similarly, one phoneme can have many different realizations depending on the same factors. Matching portions of the acoustic signal to phonemes is a crucial part of interpreting the speech signal, but it is a challenge that computational linguists have struggled to teach machines to meet. Although humans accomplish this task effortlessly most of the time, describing the way we accomplish it is an ongoing research problem in several fields. What is known about the process suggests that using contextual information is one of the ways in which people solve the problem. In

Chapter 1, I discuss some of the types of contextual information that have been shown to affect language processing.

If our strategies for speech perception and sentence processing are an optimal solution to the problem of language comprehension given the constraints of our cognitive capacities, then listeners should use all available sources of information that increase the predictability of possible outcomes of the process. This view of language understanding suggests that the comprehension process may be influenced by any factors that bear a statistical relationship with linguistic events. In Chapter 2, I review the discovery of statistical relationships between social characteristics of speakers and linguistic behavior, and discuss the current evidence that these characteristics of speakers may play a role in language comprehension.

Chapter 3 describes *t/d* deletion, the linguistic variable that is the test case for the studies presented in this dissertation. Because this variable is associated with the race of speakers, listeners could potentially use information about *t/d* deletion to make inferences about the race of speakers. This is the subject of the first sub-question, which concerns whether listeners are able to associate the use of a single variable with the social characteristics that condition the production of the variable. If listeners do make inferences about a speaker's race based on their use of this variable, are they basing their interpretations on correlations that they have observed between the social characteristics of speakers and features of the speech sounds they produce? Or are these associations mediated by stereotypes, which may not be grounded in correlations between social groups and specific variables? In Chapter 4, I present data from two experiments investigating the inferences that listeners make about social characteristics of speakers based on their use of sociolinguistic variables.

The second question concerns the ways in which listeners use the information available to them about the way people tend to talk to make inferences about what speakers have said in specific instances. If listeners use associations between social characteristics and linguistic production in order to make social judgments based on speech, then this information might also be available to them to make inferences in the other direction. That is, listeners could use social information to help with the problem of speech perception. In Chapter 5, I present data from a pair of experiments investigating whether listeners use social information to understand speech.

Whether using linguistic information to make inferences about speakers or using social information to make inferences about speech, listeners need to draw on their past experience with language in the social world to interpret either the social situation or the linguistic input that they are faced with in the present. Both inferential processes involve using information from one area to reduce uncertainty in another area. But what kind of mental representations do listeners have of their past experiences that could be helpful in the present? The third question of this dissertation addresses the kinds of representations listeners draw on when making inferences about speakers or speech. Many of the known effects of social information on speech perception described in Chapter 2 have been accounted for by some version of Exemplar Theory (Goldinger, 1996; Johnson, 1997), in which listeners' memories of speech events are detailed and include information about the speaker and other aspects of the context.

The details of this explanation, however, differ in many respects from study to study, especially in terms of the types of memory representations that are involved in language comprehension. In Chapter 6, I discuss how the current experiments could be accounted for by such a mechanism, and consider the types of representations that might underlie that mechanism. I also present data from two final experiments, investigating what levels of representation are involved in listeners' knowledge of the relationships between social characteristics of speakers and linguistic variables.

The cognitive processes involved in forming these representations and making use of them in understanding language and the social world are not necessarily specific to linguistic processing. While there may be some aspects of language understanding that are specific to a language *module* (Fodor, 1983), the extensive influence of non-linguistic information on language comprehension and interpretation suggests that some domain-general capacities are involved. Can the interaction of social information with linguistic variation be accounted for in domain-general terms? In Chapter 7, I propose an approach to integrating social information into a model of language comprehension that is informed by the data presented in previous chapters. Finally, I advance the conclusion that both the use of linguistic information in social cognition and the use of social information in language processing are natural consequences of our cognitive capacities for learning and inference.

Chapter 1

The Problem of Perceiving Speech

How do listeners extract meaning from a stream of spontaneous speech? While this process might seem straightforward, the nature of the auditory input available to listeners makes it very difficult:

> At first glance, the solution to the problem of how we perceive speech seems deceptively simple. If one could identify stretches of the acoustic waveform that correspond to units of perception, then the path from sound to meaning would be clear. However, this correspondence or mapping has proven extremely difficult to find, even after some forty-five years of research on the problem. (Nygaard and Pisoni, 1995)

In this chapter, I outline some of the aspects of speech that make speech perception difficult, and then discuss information, both linguistic and non-linguistic, that listeners use to make sense of the speech stream.

1.1 Challenges of Speech Perception

Several factors interfere with correspondence between acoustic units and perceptual units. First, there is a many-to-one mapping between acoustic cues and properties of speech sounds. The same cue, such as vowel length, can indicate multiple different perceptual properties, such as vowel identity, syllable stress, or the voicing status of the following consonant (Klatt, 1976). This many-to-one relationship also goes in the opposite direction, from speech sounds to cues: the voicing status of a consonant in English is indicated by voice onset time, the ratio of the duration of the preceding vowel to the duration of the consonant, and the presence or absence of an aspiration burst after the consonant. Because of this lack of direct correspondence between speech sounds and phonemes, there is no way to directly 'translate' between properties of the speech signal and categories of sounds.

Second, speech sounds are not discrete. The coarticulation of speech sounds is an unavoidable consequence of using our articulators in an efficient manner. Rather than proceeding one after another in a temporal sequence, speech sounds overlap with one another in the acoustic signal. Information about consonants often resides in coarticulatory effects on the formants of neighboring vowels. Without this information, the identification of consonants becomes more difficult (Fowler, 1980). Because information about multiple sounds can be overlapping, it is impossible to segment the speech signal into a linear sequence of phonemes, although such a linear sequence of phonemes is presumably what constitutes our underlying representation of words (i.e., the lexeme that the listener must try to map the speech signal onto).

Third, and most relevant to the current study, is the problem of variation (sometimes referred to as the *lack of invariance* problem) in the production of speech sounds, both within and between speakers. Very different acoustic signals can be produced to communicate the same message at different times or by different speakers. For example, the phoneme /t/ in English can be produced as a full apico-alveolar closure and a strong release burst, a glottalized closure with no release, a voiced flap, or it can be realized with no closure at all, cued by a shift in the formant frequencies of surrounding vowels or the frication noise of surrounding fricatives. Even within these categories of realization, there is great variability, such that the same utterance repeated multiple times will never result in acoustically identical productions.

In addition to variability in the realization of a sound that occurs within the productions of one speaker, different speakers can produce consistently different acoustic signals for the same phoneme. For example, one of the sources of speaker-based variation in acoustic production of vowels is vocal tract size. Men, women and children generally have different vocal tract lengths from each other, with length also varying within each group. Because of this, women and men tend to have different vowel spaces, which means that for the same vowel phoneme, women, on average, produce systematically different formant values from those produced by men (Hillenbrand, Getty, Clark, and Wheeler, 1995; Jacewicz, Fox, and Salmons, 2007; Peterson and Barney, 1952). Children also have systematically higher vowel spaces, and the space of individual variation within these demographic categories is very large. The task of the listener, upon hearing a speaker utter a syllable containing a vowel, is to interpret the frequencies of the vowel’s formants differently for each speaker, such that very different values are interpreted as indicating the same phoneme, and the same values are interpreted as indicating different phonemes. Faced with the problem of how listeners perceive a consistent set of vowels between these groups given their vastly different acoustic properties, speech perception researchers have suggested that listeners normalize the acoustic properties of vowels based on ratios of formants, rather than absolute formant values (Strange, 1999; Syrdal and Gopal, 1986).

It is remarkable that the variability in speech production does not prevent listeners from construing all of the different acoustic signals that they observe as members of a consistent set of categories. While specifically normalizing based on formant ratios applies only to vowels and vocal tract size, the idea that listeners can apply different sets of criteria for classification of speech sounds in different situations is a necessary part of explaining their ability to understand two different speakers’ productions of the same word as representing the same meaning. But developing and applying these different sets of criteria is a challenging task. One of the things that make it possible is the highly structured nature of the variation listeners are faced with. Although different tokens of the same type of sound can be produced very differently, the differences are not randomly distributed. On the contrary, Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, and Gildea, (2003) found that a number of factors including the position in an utterance, the speech rate, the presence of disfluencies, and the predictability of the word it is part of all constrain the duration of the sound in systematic ways. The systematicity of these relationships means that if listeners store information about these factors (and many others), this stored information can help the listener to make meaning out of the acoustic signal.

Storing information about regularities in the correspondence between contextual factors (both linguistic and non-linguistic) and linguistic events is helpful to the listener because more expected events are easier to process (Hale, 2001; Levy, 2008). For example, words with higher transitional probabilities (that are more likely and thus more expected) require shorter reading times (McDonald and Shillcock, 2003). Although there is a great deal of uncertainty about what types of information listeners may and may not be using for this purpose, speech perception and sentence processing research have provided many examples of factors that seem to influence language comprehension.

1.2 Types of Information Used in Speech Perception

There is a large body of evidence suggesting that the percept of speech and the interpretation of meaning that is extracted from that speech are the product of many aspects of the auditory input, and of the non-linguistic context of the speech signal. Listeners make inferences that help them to categorize speech sounds and interpret strings of speech based on other aspects of the linguistic context including information about word meaning (Warren, 1970) and affect (Nygaard and Lunders, 2002). Listeners also use aspects of the non-linguistic context such as visual information about articulation (McGurk and MacDonald, 1976) and objects in a scene (Trueswell, Sekerina, Hill, and Logrip, 1999) to constrain their interpretations of the speech stream. The next two sections describe the types of information that influence language comprehension.

1.2.1 Use of information from the speech stream

Information from the speech stream can help listeners resolve ambiguity they encounter in understanding language. For example, listeners use the semantic context of a sound to help restore missing information. This effect is called *phoneme restoration* (Warren, 1970; Warren and Warren, 1970). Warren and colleagues played participants different sentences all containing one word with the first phoneme replaced by a burst of noise. When listeners heard this burst of noise followed by the string *eel*, they interpreted this as one of the words *wheel, heel, peel,* or *meal* depending on the last word in the sentence, which made only one of those words sensible (and thus predictable) in context. Interestingly, listeners did not consciously infer that this word must have been intended by the speaker; instead, they actually perceived the word that made sense in the utterance. The restoration effect was so strong that listeners were not even aware that a speech sound was missing, and could not identify the missing sound or the place in the sentence where the extraneous noise occurred. In further research, Warren has identified non-speech examples of similar effects, leading him to describe phoneme restoration as a subset of the more general phenomenon of *auditory induction* (Warren, Obusek, and Ackroff, 1972).

Affective information conveyed by a speaker's tone of voice has also been shown to have an impact on speech perception. Nygaard and Lunders (2002) found that hearing a word in either a happy or sad tone of voice influenced listeners' interpretations of the meanings of ambiguous words. Listeners heard a word that could have been either of two homophones, one of which had an emotionally valenced meaning and the other of which had a neutral meaning, such as a word that is ambiguous between *flower,* which has a positive meaning, and *flour*, which has a neutral meaning (neither positive nor negative). Listeners chose the emotional meaning (in this case, the happy word *flower*) more often when it was said in a congruent emotional tone of voice (a happy tone of voice), and they chose the neutral meaning (in this case, the word *flour*) more often when it was said in a neutral tone of voice. The authors suggest that these results are compatible with an exemplar-based view of the lexicon, in which aspects of the speech situation like emotional tone of voice are encoded along with the lexical item and retrieved along with it when it is heard. On this view, happy words are associated with happy tone of voice via frequency of co-occurrence, and this association accounts for their results. Alternatively, as they point out, their effect may arise because words with happier connotations are actually more felicitous when spoken in a happy tone of voice than neutral words, which would make this effect semantic/pragmatic in nature. Whatever the nature of the effect, it demonstrates that information from tone of voice is a part of very early language comprehension, helping listeners to select a lexeme to map onto the speech signal.

Prior linguistic experience can also influence phone perception. Clarke (2003) showed that listeners tune their expectations for phone categories to information about speakers that they gather from prior exposure to the speaker. Listeners were exposed to a speaker with a foreign accent in which Voice Onset Time (VOT) for voiceless consonants was shorter than in the listeners' native language (English). After exposure, listeners showed different expectations for VOT: they categorized phones as voiced and voiceless according to a different standard for VOT than they had used prior to exposure. Listeners who had heard the foreign accented speech had a lower VOT threshold for voicelessness than those who had not heard foreign accented speech.

There are at least two ways in which prior experience with a speaker could influence phone categorization. Listeners might show this effect because they have categorized the speaker they heard as a foreigner and are using this information about the speaker to adjust their perception by using a different perceptual boundary for foreign speakers than they use for native speakers. By this account, the foreignness of the speaker is not dissimilar from a social characteristic, based on which listeners can infer things about speech. An alternative account of this effect is that it might be an automatic effect of *perceptual learning* (Goldstone, 1998), a process by which listeners can improve the efficiency of how they interact in the world by adjusting the way they categorize information based on their recent experience. In the case of Clarke's results, the act of categorizing sounds with different VOTs as voiceless sounds may have at least temporarily changed listeners' phone boundaries without the need for direct inferences about the speaker. In a follow-up study, Clarke-Davidson, Luce, and Sawusch (2008) suggest that changing performance on phone discrimination tasks following exposure to speech does represent true perceptual learning, rather than a decision bias based on the task.

Information from other levels of linguistic description can also influence the perceptual learning of speech sounds (Norris, McQueen, and Cutler, 2003). That is, by influencing the way listeners categorize sounds when they hear them, context can influence the perception of similar sounds in the future. In this study, the authors gave Dutch speakers experience hearing a sound that was ambiguous between /s/ and /f/ in two different lexical contexts. These contexts provided top-down information indicating that the ambiguous segment was either an /s/ or an /f/, because they formed words according to one interpretation and non-words according to the other. Listeners who were exposed to the ambiguous sound in /s/-contexts subsequently categorized a larger proportion of sounds on the /s/-/f/ continuum as /s/, and listeners who were exposed to the ambiguous sound in /f/-contexts subsequently categorized a larger proportion of these sounds as /f/, indicating that lexical information from their prior exposure to the ambiguous sound had changed their /s/-/f/ category boundary.

1.2.2 Use of information from outside the speech stream

In addition to effects of information encoded in the auditory information like those described above, there have been many documented effects of non-auditory information on interpretation of the speech stream. One of the earliest examples of non-linguistic information affecting language comprehension is the McGurk effect (McGurk and MacDonald, 1976), in which articulatory information from the way a speaker's mouth moves conflicts with the acoustic information in the speech stream, causing listeners to believe they have heard an intermediate sound. For example, watching a video of a person saying the syllable [ga] while listening to the person saying the syllable [ba] results in the percept of having heard the syllable [da], which is intermediate between [ga] and [ba] in place of articulation. This effect is robust even when the same stimuli are repeated for the same listener and does not go away when the listener is aware of the illusion; even having heard the sound without looking at the articulators and recognizing that it is [ba] does not prevent listeners from hearing [da] again when they open their eyes. Thus, listeners are not only *able* to use visual information about the articulators when discriminating phones, but they cannot help doing so. This evidence of the use of visual information in speech perception was among the first suggestions that the language faculty automatically uses non-auditory information when processing an auditory speech signal. The existence of this effect raised the possibility that other, similar kinds of non-linguistic information could be used in this process.

Another kind of visually presented information that can influence how listeners interpret an utterance is its referential context. When listeners hear syntactically ambiguous material like *Put the frog on the napkin…,* the items in the context can change the interpretation of the ambiguous material. For example, if there is a frog that is sitting on a napkin and a frog that is not sitting on a napkin, the ambiguous phrase is interpreted as a modifier of a potential referent (telling you which frog). If there is only one frog, it is interpreted as a potential goal location for an action (telling you where to put the frog) (Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy, 1995; Trueswell, Sekerina, Hill, and Logrip, 1999). This effect is essentially pragmatic in nature, implying that listeners assume that speakers will follow Grice's Maxim of Quantity[1]. In this case, that means that they will provide information about which frog is intended when there is more than one frog, and will not provide unnecessary descriptions of the frog when there is only one. In natural conversation, the ambiguity that is exploited in these studies may arise very rarely, because intonation often disambiguates the structures (Kraljic and Brennan, 2005). However, even when natural intonation information was available, eye-tracking data from Kraljic and Brennan's study indicated that listeners might have been considering the alternative interpretation up to 60% of the time, suggesting that information from the referential context could play a supporting role in helping listeners interpret ambiguous utterances of this type.

Visually presented information can also influence listeners' explicit judgments and attitudes. Podol and Salvia (1976) showed that visually presented information about physical characteristics of speakers can influence how listeners perceive speech. They showed speech pathology students pictures of speakers who either had a visible facial disfigurement or whose disfigurement had been eliminated by photo retouching. They then asked the students to evaluate the each speaker's need for speech therapy.

[1] Grice's Maxim of Quantity: Make your contribution as informative as is required for the current purposes of the exchange. Do not make your contribution more informative than is required (Grice, 1975).

Participants rated the visibly disfigured speakers to be more in need of speech therapy than the speakers whose disfigurement had been eliminated, even when the speech stream they heard was held constant.

1.3 Summary

Aspects of both the linguistic and the non-linguistic context influence language comprehension and evaluation. What these varied aspects of the context have in common that makes them useful to listeners is that they are all statistically related to linguistic events, and thus can help make those events more predictable. This suggests that listeners could use any statistical regularity in the input to help them understand language. The next chapter discusses a large category of contextual factors that are statistically correlated with linguistic behavior that could potentially be used in the same way: social characteristics of speakers.

Chapter 2

Sociolinguistic Variation as a Statistical Regularity

The way different speakers express the same ideas or even produce the same words and sounds can be surprisingly different. As discussed in the previous chapter, spontaneous speech is notoriously variable, and speakers have many choices of phonetic realizations, lexemes, syntactic constructions, and even discourse patterns when formulating their utterances. Many of these choices do not influence the propositional meaning of the utterance, and it is difficult at first to see how speakers choose among these available options for expressing their ideas. In an early study of variability in the English of the residents of New York, Hubbell (1950) concluded that the pattern governing the pronunciation of New Yorkers "might most accurately be described as the complete absence of any pattern" (48). The belief that the different ways that are available for speakers to say things are in *free variation* (that is, that there is no pattern to the choices speakers make) was widely held until Labov undertook his enormously influential survey study of New York City in 1963. In the course of this study, Labov demonstrated that the variable linguistic behavior observed by Hubbell was actually structured by many factors, including both the social characteristics of the speakers and the style in which they were speaking (1966). Labov's work was the foundation of a tradition of sociolinguistic studies of variation, which has evolved from survey studies modeled after his study of New York City to more ethnographic investigations of linguistic variation in many different speech communities. In the first part of this chapter, I discuss some of the seminal studies of sociolinguistic variation that demonstrated that variable linguistic behavior varies systematically with social characteristics of speakers. The second part of the chapter addresses what is known about how listeners respond to this variation.

2.1 Sociolinguistic Survey Studies and Beyond

The tradition of the sociolinguistic survey study emerged from Labov's study of New York City. Labov began his study by constructing a stratified random sample of the adult residents of the Lower East Side of New York, selecting participants who were native speakers of English and had not moved for two years, and including speakers of a variety of ethnicities, ages, and social classes, as well as both genders. He and an associate then interviewed 157 adults, and 58 of their children, using an interview protocol that has been developed into the standard procedure for sociolinguistic survey research. This protocol involved bringing up various pre-set topics (such as childhood games) for the subject to talk about, and interspersing some specific tasks like reading word lists containing minimal pairs of words[2], reading lists of isolated words, and reading a passage aloud. These different parts of the interview elicited what have been referred to as different *styles*, which correspond to different amounts of attention that the subject was paying to their own speech. Styles in which a lot of attention was paid to speech are known as *careful*, and styles in which little attention was paid to speech are known as *casual*.

The style consisting of participants' speech in response to the interview topics brought up by the researchers is known as *interview style* (IS). This is the elicited style in which participants paid least attention to their speech, although Labov sometimes encountered even more casual speech when participants spoke outside the interview setting while the tape recorder was on; this type of speech is referred to as *casual style* (CS). The reading passage speech (RP) is more careful than these styles, with the isolated word list (WL) being even more careful, and the list of minimal pairs (MP) being the most careful style elicited in the interviews.

Labov analyzed the speech he collected in the interviews for the use of features that were known from previous research to vary in the speech of New Yorkers. These features are known as *sociolinguistic variables*, which are the basic units of variationist research. Fasold (1990) has defined a sociolinguistic variable as "a set of alternative ways of saying the same thing, although the alternatives will have social

---

[2] A *miminal pair* is a pair of words that vary in terms of only one phoneme; for example, *hat* and *hot* are a minimal pair, because they differ only in their vowel sounds, and contain the same consonant sounds.

significance" (p.223-24), a definition which is faithful to the original Labovian principles that established the concept.

One of the variables Labov investigated was (r), which he analyzes as having two variants: /r/ can be phonetically present or deleted. Labov discovered that the use of these variants is probabilistically conditioned by a variety of linguistic and non-linguistic factors. Among the factors that he found to condition variability in the use of this variable were the style of the speech and the social class of the speakers. Table 1, adapted from Labov (1966), shows the styles of interview speech, organized from most careful (MP) to most casual (CS), along the left, and the social class of the speakers, from lower class (LC) to working class (WC) to middle class (MC), along the top. The numbers represent an aggregation of the (r) indices of the speakers in each of these social class groups for each style of speech. The (r) index is simply the percentage of utterances that could have contained an /r/ in which the /r/ was present.

Table 2.1 shows two striking correlations with the realization of the (r) variable: first, all speakers produce more /r/, the more standard variant of the variable, in more careful speech styles. Additionally, speakers who are in a higher social class produce more /r/ than speakers in a lower social class, across all speech styles. The robustness of these two correlations highlights the extent to which Hubbell's (1950) assessment of the state of variation in New York was incorrect; in fact, the pattern governing the pronunciation of New Yorkers was describable in quite specific terms.

	Class		
Style	**LC**	**WC**	**MC**
MP	50.5	45	30
WL	76.5	65	44.5
RP	85.5	79	71
IS	89.5	87.5	75
CS	97.5	96	87.5

Table 2.1. (r) indices aggregated over speakers, by social class and speech style. Adapted from Labov (1966).

The (r) indices reported in Table 2.1 represent, in essence, probabilities that a group of speakers in a certain style will produce the deleted variant of (r). These probabilities are calculated by counting the number of times an /r/ is deleted, and comparing that number to the number of times an /r/ had the potential to be deleted. Counting the number of times an /r/ had the potential to be deleted constitutes defining the *envelope of variation* for this variable (see Labov, 2008 for a discussion of this construct). The process of defining a variable requires defining the envelope of variation. In the case of (r), it is not as simple as looking at any word with a phonemic /r/. The /r/ may only be deleted in certain linguistic environments, such as after a vowel; however, once it has been determined where it is possible for speakers to delete the /r/, the envelope of variation has been defined. For other kinds of variables, defining the envelope of variation can be more difficult, because the division between which things people choose to say and how they choose to say them is not always entirely straightforward. Especially for variables involving discourse- or syntax-level choices, the notion that the alternatives mean the same thing may be difficult to support (Lavandera, 1978).

And, as Eckert points out, variability in the way people say things and variability in what they choose to say may both have the same root cause: “Different ways of saying things are intended to signal different ways of being, which includes different potential things to say” (2008, p. 3).

Although the idea of identifying “alternative ways of saying the same thing” may sometimes be problematic in the case of a specific variable, it is to a large extent a necessary component of engaging in variationist research. In order to quantify the linguistic behavior of speakers, it is necessary to establish the boundaries of the envelope of variation. Determining these boundaries requires making decisions about which linguistic behaviors one will consider ‘equivalent’ for the purposes of the study – to establish the “closed set” Labov refers to above. Accepting that this equivalence is to some degree imposed by the observer does not, however, make it impossible to engage in variationist sociolinguistics in these cases. It simply requires that researchers operate under assumptions that they know not to be entirely true – much like assuming a frictionless world in studying Newtonian mechanics.

The findings of Labov’s large-scale survey study of New York City sparked a movement to investigate what is now called *sociolinguistic variation*: variability in the way people speak, at any level of linguistic description, that is conditioned by social factors. The variationist project has been documenting and describing the way different groups of people produce language, and studying the factors that condition variation.

Research in the variationist tradition, following and expanding on Labov's methodology, has uncovered a vast and complex system of social conditioning of linguistic variation. Survey studies from around the globe have uncovered a few social factors that are frequently correlated with the use of sociolinguistic variables, including the age, gender, ethnicity, and social class of a speaker, among others (Macaulay, 1978; Trudgill, 1974; Wolfram, 1969). These social factors are not always organized the same way in different communities; however, in all of these studies, social characteristics of speakers, defined according to some locally or globally determined criteria, are strongly correlated with variable linguistic behavior.

Phonological variation can also be conditioned by far more subtle and locally-determined social categories (Labov, 1973). In his ground-breaking study of Martha's Vineyard, Labov determined that pronunciation of the diphthongs [aw] and [ay] varied according to a combination of attitudinal features and demographic features, which he discovered during his fieldwork on the island. Speakers had varied attitudes toward traditional island life and the encroachment of mainlanders, and these attitudes influenced their pronunciation of the diphthongs. Islanders who oriented toward the mainland had lowering nuclei in their diphthongs, consistent with the mainland sound change, whereas islanders who rejected mainland ideals and oriented toward local values reversed this change, emphasizing a vernacular, local variant (despite sometimes being young and not belonging to the fishing community, characteristics that might otherwise have steered them towards the globally defined standard). This pattern was one of the first indications that speakers orient toward a set of norms and values that they identify with, rather than simply reproducing the speech patterns of others who are similar to them on various social dimensions.

Further investigations of sociolinguistic variation have discovered that even when linguistic variation reflects social structures like social class that are well-known to correlate with language use, it may be locally defined versions of these social structures that are relevant in a community, which may be uncovered through ethnographic fieldwork. For example, Rickford (1979, 1986) studied an East Indian sugar estate community he called "Cane Walk," Guyana, in which language use correlated with membership in two broad social classes, which he referred to as the Estate Class and the Non-Estate Class. Members of the Estate Class were fieldworkers on the sugar estate, while the members of the Non-Estate Class had other jobs, which, while they varied in status, were all more statusful than working in the sugar fields. Despite the fact that it would have been possible to form a more articulated class structure in this community, this two-way division correlated strongly with linguistic behavior, with members of the Estate Class using overwhelmingly Creole forms, and the members of the Non-Estate Class using forms that were closer to Standard English. Rickford analyzes these different patterns of language use as emerging from a social situation in which members of the Estate Class do not have many opportunities for socioeconomic advancement, and use Creole as a way of expressing their solidarity with their social class and their opposition to the social structure that deprives them of these opportunities for advancement. The members of the Non-Estate Class, on the other hand, have some opportunities for incremental advancement in socioeconomic status, and may use more standard language forms as a strategy for achieving upward mobility.

Locally-defined groupings have also proven useful when investigating variation among adolescents, who both conform to and depart from adult categorization schemes. Eckert (1989a, 2000), described a Detroit area high school as containing two major social groups: the Jocks and the Burnouts. These two groups oriented around different sets of norms and values, with the Jocks orienting towards institutional successes in athletics and student government, while the Burnouts oriented towards the values of the local urban center (Detroit). The members of these two groups indexed their identification with their communities through dress, social practices, and linguistic behavior. The linguistic variation Eckert observed centered around the Northern Cities Shift, a set of vowel changes taking place in American urban centers near the Great Lakes (Labov, Yaeger and Steiner, 1972). This chain shift is characterized by advanced low vowels and retracted mid vowels, such that the word *on* sounds more like *Ann*, *fun* sounds more like *fawn*, etc. (see Figure 2.1 for a diagram of the change).

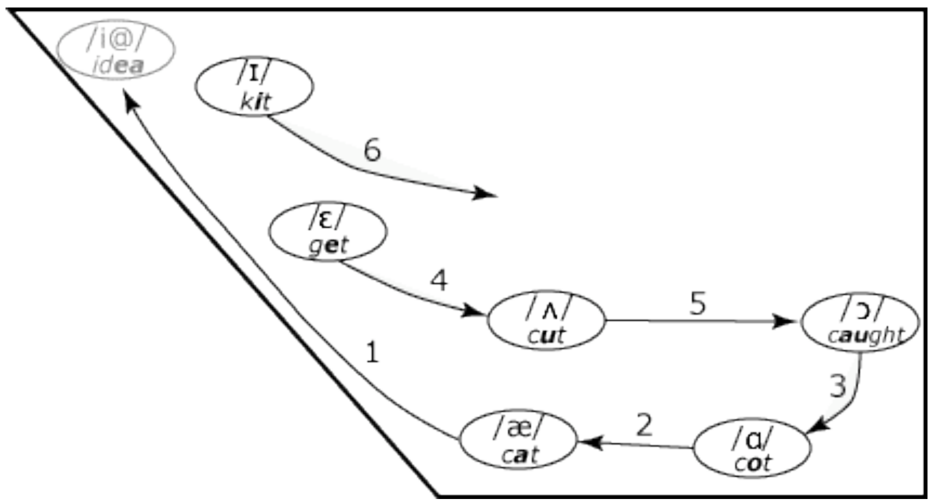

Figure 2.2 The Northern Cities Chain Shift (Reproduced from Labov, 1996).

The less established aspects of the Northern Cities Chain Shift were associated with an urban orientation, and consequently Burnouts showed more advanced realizations of these vowels than their Jock counterparts. The more established components of the shift were stable in the suburban areas as well as the urban areas; these variants were most advanced among females, both Jock and Burnout (although the Burnout girls led in their use). The Jock and Burnout groups correspond in some ways to social constructs in the adult world (such as socioeconomic classes), and this correspondence is part of how the adolescents understand their roles in both the school and society; however, classifying students according to which locally-defined social group they align with more in their social practices is provides greater predictive power than relying solely on externally-derived social categories and imposing them on the high school social structure. The high school students' affiliations with these groups reproduces the adult social hierarchies, but is not strictly defined by their parents' socioeconomic status; the differences among the groups are primarily symbolic. This highlights the degree to which linguistic variation reflects *identity*, which Eckert (2000) describes as including “a person’s place in relation to other people, a person’s perspective on the rest of the world, a person’s understanding of his or her value to others,” which is negotiated in the context of group identities. Although it might seem possible to view features of a person’s identity as merely correlated with linguistic variation, Eckert champions a *constructivist* view of variation in which linguistic variation is not just correlated with the factors that define an individual’s identity, but that linguistic variation is one of these factors itself.

While the idea that linguistic variation might reflect and perhaps partially constitute identity among adolescents may seem to depend on the fluidity of adolescent social organization, groupings among adults that share variable linguistic behaviors may also organize themselves primarily around aspects of identity. Linguistic behaviors can even be instrumental in constructing new social categories. Zhang (2005) found that Chinese professionals working in foreign-owned businesses (Beijing yuppies) used some features of non-mainland Mandarin (such as full tone), which are viewed as cosmopolitan, more than employees working state-owned businesses. In addition, the employees working in state-owned businesses used more local Beijing variants than the foreign-owned business employees, who more closely approximated Mainland Standard Mandarin in their use of the same variables. Zhang argues that the employees of foreign-owned businesses were not simply producing more standard language than the state-owned business employees; rather, they chose to depart from the official standard in favor of non-mainland varieties, rather than local varieties, in order to construct an urban yuppie identity, identifying themselves with the global context in which they worked.

The relationships between locally defined categories and groups and linguistic variation suggests a meaningful relationship between speakers and the types of linguistic variants they produce. If linguistic variation were due simply to speakers automatically mimicking the people they speak to most, why would the social groupings that linguistic behavior is organized around vary so much from place to place? Although speakers are not often conscious of their own socially conditioned variable linguistic behavior, the relationships between social categories and linguistic variation do not seem to be arbitrary; rather, this variation has been described by sociolinguists as being organized such that it conveys *social meaning*. The idea that socially conditioned linguistic variation is meaningful is also partly supported by the fact that these linguistic variables are used differently not just by different speakers (social variation), but at different times by the same speaker (stylistic variation). The study of stylistic variation has expanded from Labov's *attention paid to speech* paradigm to include variation based on the speaker's beliefs about and attitudes toward his or her audience (Bell, 1984; Giles and Coupland, 1991), and variation highlighting aspects of the speaker's social identity (Eckert, 2000; Podesva, 2007).

For example, Bell (1984) found that news announcers working for two different radio stations in New Zealand used sociolinguistic variables, such as intervocalic /t/ voicing (pronouncing the word *writer* with a flap so that it sounds more like *rider*) differently on the two stations, which had different audiences, despite being the same individuals and discussing generally the same topics. Though it is not impossible that the announcers were paying more attention to their speech in newscasts on one station than the other, the similarities in the speech situations and topics suggest that if this is the case, it is in direct response to the audience that such a difference arises. Bell interprets this stylistic variation on the part of the newscasters as an indication that their speech is in some sense *designed* for their audience, a theory that considers stylistic variation to be a less automatic and more active behavior on the part of the speaker than the Labovian notion of style might suggest.

In addition to tuning their performances to their audiences, speakers also use linguistic variables to highlight different aspects of their identities in different situations. Podesva (2007) describes the use of phonation type as a resource for constructing different personae in different scenarios by a gay medical student. “Heath,” the subject of this study, used falsetto and creaky voice differently in different scenarios; at a barbecue with his friends, he produced longer stretches of falsetto (and longer stretches of creaky voice), thereby widening his f0 range and constructing an expressive ‘diva’ persona, which differed from the personas he adopted in other situations, such as in a telephone call with his father, and with a patient in a clinical setting. Although this ‘diva’ persona is a distinctly gay persona, calling his falsetto use an index of gayness misses the additional details Heath provided about what type of gay persona he embodies, and misinterprets his behavior as indicative of a general characteristic. In fact, this linguistic behavior indicated a persona he was adopting at a specific moment in time (at the barbecue, with his friends). Thus, Heath’s variable linguistic behavior was meaningful not just about stable aspects of his identity, but about which aspects of his identity were salient in certain situations.

2.2 Social Information and the Listener

The discovery that linguistic variation is structured by a wide variety of social and stylistic factors launched the subfield of sociolinguistics, and demonstrated that the systematic study of variable linguistic behavior is both possible and fruitful. For the purposes of this dissertation, however, the importance of this discovery (and the work that has followed it) is that it uncovered a vast system of statistical regularity between language and contextual factors that could potentially be exploited by listeners for the purpose of making inferences about both language and context.

Focusing on the types of inferences that listeners could make based on this systematicity requires considering sociolinguistic variation from the perspective of the listener, a perspective that has not been considered nearly as often as that of the speaker. However, the role of the listener is also important in explaining the systematicity of sociolinguistic variation. The presupposition of ascribing meaning to this variable production is that it involves the transmission of information from speaker to listener. Information transmission requires a listener who is able to receive the information and interpret its meaning. If listeners are not at least subconsciously aware of sociolinguistic variation, sociolinguists are faced with a difficult problem: how can speaker/listeners develop such consistent and intricate patterns of sociolinguistically conditioned production if they are not influenced by each other's linguistic behavior? Theories of the social meaning of linguistic variables rely crucially on the role of the listener.

It is important to note that focusing on the role of the listener does not necessarily imply that the listener figures significantly into the speaker's cognitive model. While there is evidence that speakers adjust their production of sociolinguistic variables for their interlocutors (Bell, 1984), the ways in which speakers use linguistic practices as resources for doing social work are not likely to involve explicit inferences about listener interpretations of these linguistic moves. In fact, it may be that speakers do not model the listener at all, consciously or unconsciously, in many or most situations. However much or little the listener is involved in production on the level of the individual speaker, the interpretation of this activity by speakers as meaningful depends on the listener.

The systematic and meaningful nature of linguistic variation places sociolinguistics at the intersection of the cognitive and social sciences. This behavior on the part of speakers is both automatic and deliberate, and it exists on the level of the individual as well as on the level of the community. While sociolinguistic variation is part of language as a formal system, it is also part of language as a system for social interaction; as Eckert (2008) points out, "The social is not just a set of *constraints* on variation – it is not simply a set of categories that determine what variants a speaker will use – it is a meaning-making enterprise" (p. 15). The listener plays an integral role in this multi-layered process, as a comprehender and ratifier of the meanings speakers attempt to convey with their linguistic behavior. In addition to being a formal system and a social one, language is also a cognitive system. The focus of this dissertation is on how sociolinguistic variation interacts with the rest of our cognitive capacity for producing and understanding language.

The listener's role in language as a cognitive system is primarily that of an agent of perception. However, very little is known about the perception and comprehension of variable linguistic behavior – do listeners store information about the structure linguists have observed in socially conditioned variable production? If this information is somehow monitored by listeners, how do listeners use the knowledge they accumulate?

The statistical relationship between social characteristics of speakers and their linguistic behavior could be useful to listeners in two different ways. First, knowledge about the social conditioning of linguistic variation could be helpful to listeners for making social inferences about speakers based on their speech. In addition, this knowledge could help listeners make inferences about speech based on what they know about a speaker. Both of these types of inference on the part of the listener have been investigated to some extent, in largely separate literatures.

2.2.1 Making inferences about speakers based on speech

To the extent that researchers have considered the role of the listener in sociolinguistic variation, a majority of the work has focused on *language attitudes* (for a review of this concept and work related to it, see Garrett, Coupland, and Williams, 2003). Several different techniques have been used to investigate listeners' attitudes toward aspects of language. In some ways the most straightforward method of assessing language attitudes is to ask people about their opinions of language varieties or even specific linguistic variables (Fishman, Cooper, and Ma, 1971; Shuy and Williams, 1973). While this method can yield useful and interesting results, it is not always possible to fully probe people's attitudes this way, because people are not always honest to themselves and others about their attitudes. Participants' fear of appearing bigoted or judgmental sometimes prevents them from reporting negative attitudes about the language of others. In addition, the language attitudes of even one person can be multi-layered and complex (or even contradictory), making it difficult to articulate them even if there are no other obstacles to explaining them to researchers.

Because directly questioning people is usually insufficient for investigating their language attitudes, more indirect methods have been employed in service of this goal. Sometimes speakers' self-reports about their own language production can be informative about their language attitudes; to the extent that speakers inaccurately describe their own use of linguistic variables and this inaccuracy is perceived, the differences between speakers' descriptions and their actual linguistic behavior can provide insight into the ways they believe they ought to use language, or the variants they believe are more statusful (Labov, 1994).

Another semi-indirect method of investigating language attitudes, the tradition of the Matched Guise Technique (Lambert, Hodgson, Gardner, and Fillenbaum, 1960) has been a valuable method for gathering information about language attitudes on a number of levels. This technique involves investigating attitudes toward the same speaker in different "guises," which correspond to the linguistic units of interest (for example, language variety). Lambert et al. (1960) recorded balanced French-English bilinguals speaking both English and French, and asked participants to rate the speaker of each clip on various personal attributes, without telling them that some of the clips were spoken by the same speaker in different languages. They found that participants rated the same speaker more highly on several positive attributes in their English guise compared to their French guise, and vice versa, indicating that they associated certain positive traits with English, and others with French. Because the actual speakers (and their voices, etc.) were identical, the differences in evaluation that their guises elicited are interpreted as reflecting participants' implicit attitudes toward the language varieties.

Following Lambert et al. (1960), the Matched Guise Technique has been used to compare many sets of completely separate languages (e.g., Wolck, 1973; Sridhara, 1984; Woolard and Gahng, 1990). It has also been used to compare reactions to two or more mutually intelligible dialects (e.g., Cheyne, 1970; Creber and Giles, 1983; Giles, Coupland, Henwood, Harriman, and Coupland, 1990; Strongman and Woosley, 1967). The study of American dialects associated with race or ethnicity has also employed this technique (e.g., Fraser, 1973; Johnson and Buttny, 1982; Purnell, Idsardi, and Baugh, 1996; White, et al., 1998). Campbell-Kibler (2005) provides an extensive discussion of varieties and languages that have been studied using the matched guise paradigm and variations thereof, from the 1960s onward. While this technique has been adapted to investigate aspects of language at many levels, much of the work on language attitudes has been on the subject of listeners' attitudes toward whole languages or language varieties.

Traditional language attitudes studies can be illuminating on the subject of how listeners feel about social groups that they interact with, because although attitudes about language varieties and the groups that speak them are strongly related, participants will often be willing to express attitudes about language varieties that they would not be willing to express about a group of people. While this is a strength of the technique, the disadvantage of this relationship between speaker and variety is that most investigations of language attitudes cannot separate listeners' reactions to a variety from their reactions to those who speak that variety. To overcome this disadvantage, several other methodologies for investigating language attitudes have been employed to investigate the reactions listeners have to smaller units of linguistic variation.

The field of perceptual dialectology is engaged in investigating listeners' perceptions of dialects and dialect boundaries. Some studies in this tradition have explored the role of individual (usually vocalic) variables in forming listeners' impressions of a speaker's regional origin. Plichta and Preston (2005) investigated the influence of /ay/-monophthongization on listeners' perceptions of Southernness. They created a 7-step continuum between diphthong and monophthong /ay/ and asked listeners to assign each token to one of nine locations on a map, ranging from north to south near the Mississippi River. Although listeners did not report much confidence in their judgments, Plichta and Preston found that listeners on average assigned more monophthongal productions of /ay/ to locations that were farther south on the continuum.

In one of the earliest experiments investigating listener reactions to sociolinguistic variables, Labov (1966) played samples of speech containing five socially stratified phonological variables to listeners in New York, who evaluated the probable occupation of the speaker of each sample. The possible occupations fell along a continuum of engagement in the standard language market, with occupations like “television personality” requiring high engagement in the standard language market (Bourdieu, 1977; Bourdieu and Passeron, 1977), and occupations like “factory worker” that did not require engagement in the market. The speakers’ use of the socially stratified phonological variables strongly predicted the occupations listeners assigned to them, with speakers using statusful variants of the phonological variables being assigned to occupations requiring high engagement in the standard language market. Labov’s 1966 results indicate that listeners make inferences about speakers’ engagement in the standard language market based on their linguistic choices. But do listeners also make inferences about more inherent qualities of speakers’ identities based on their use of sociolinguistic variables?

Campbell-Kibler (2007) investigated the effects of the sociolinguistic variable (ING) (e.g. *walkin’* vs. *walking*) on listeners’ attitudes. Her findings suggest that listeners do make use of linguistic variation on the level of a single variable to make judgments about speaker characteristics. In this modification of the matched guise paradigm, instead of having the same speaker produce two different guises, the guises were created electronically out of identical recordings. The critical feature (in this case, whether a word ended in a velar nasal or an alveolar one) was digitally manipulated to create pairs of minimally different speech samples, which were then evaluated by naïve listeners. Manipulating the realization of the final nasals in (ING) influenced listeners' judgments about the person who used it and the speech situation in which it occurred. For example, the alveolar nasal /n/ made speech sound more casual while the velar nasal /ŋ/ made speech sound more formal, and the alveolar nasal made speakers sound less educated and articulate, whereas the velar nasal had the opposite effect (Campbell-Kibler, 2007).

In addition to these effects, which were constant across the speakers and speech styles involved in this experiment, Campbell-Kibler describes several effects that interact with speaker dialect region. For example, use of the alveolar nasal increased listeners' descriptions of a speaker as "accented," but only if listeners perceived the speaker as Southern. In the case of the one speaker who was perceived as having a New England accent, it was the velar nasal that increased listeners' ratings of his accentedness. This interaction suggests that listeners' perceptions of "accentedness" were relative to either their own production or some unmentioned standard, from which the New England-sounding speaker and the Southern-sounding speakers departed in different directions. Thus in both cases, their usages of (ING) could be departures from the standard to which they were being compared, even though they were not similar. Listeners used the realization of (ING) in the context of other facts about speakers and their speech to make attitude judgments about speakers, and they had explicit and implicit beliefs about who uses this variable and what it means about them.

I have so far approached the question of how aspects of speech can influence listeners' attitudes about speakers from the perspective that listeners may or may not use information about correlations in the world to make inferences about speakers. The way speakers talk could potentially influence listeners' attitudes about them without directly reflecting correlations in the world. That is, listeners might have stereotypes about how speech relates to speaker characteristics that do not accurately reflect how different types of speakers speak. For example, Americans sometimes interpret features of British accents, such as non-rhoticity, as upper-class, even if these features are also characteristic of working-class British accents and do not by themselves carry an upper-class meaning (Lippi-Green, 1997, p. 98). This interpretation may be based on a stereotype of British speakers as upper-class that has been generalized to all British speakers, regardless of their actual socioeconomic status, perhaps due to limited exposure on the part of Americans to the full spectrum of British society. Whatever the source of this association, it does not correspond to the actual relationship between these features and social characteristics of British speakers. Nonetheless, it may constitute a very real factor in how American listeners respond to British accents. Thus, identifying correlations in the world between linguistic behavior and social characteristics of speakers does not guarantee that listeners will associate the behavior with those characteristics, either consciously or unconsciously. Likewise, identifying associations in listeners' perceptions between linguistic features and social characteristics is not in itself evidence that listeners store information about these correlations in their experience. Experiment 2 in the next chapter addresses this issue.

2.2.2 Making inferences about speech based on speakers

The same relationships that allow listeners to make inferences about what kind of person they are hearing based on their speech, as evidenced by language attitudes studies, could be used to make the opposite kind of inference; listeners could use these relationships to make inferences about what someone has said based on their social characteristics. There are several findings from the sociolinguistics and speech perception literature suggesting that this may occur.

Social information from visual displays can influence listeners' comprehension and evaluation of language. Observers have been shown to use information from the visual signal when making subjective judgments about aspects of speech such as how strong a foreign accent a speaker has (Rubin, 1992; Rubin and Smith, 1990). In Rubin and Smith (1990), participants saw a picture of either a White or an Asian female that they were told was the speaker of the clip they heard. In both conditions, they heard a clip of a native English speaker lecturing on a topic either in the humanities or the physical sciences. The ethnicity of the pictured lecturer significantly affected how accented listeners perceived them to be, even when the auditory signal was held constant (and contained no cues to a foreign accent). The Asian speaker was perceived to have a greater degree of foreign accent than the White speaker, despite the fact that listeners were hearing identical recordings of speech.

Williams et al. (1972) also found an effect of teachers' language attitudes on judgments of confidence, language background, and expectations on performance in the classroom. Teachers' stereotyped beliefs about the way students talked were significant predictors of their ratings of the "ethnicity-nonstandardness" of the speech from videotapes of new students, and their ratings on this dimension were significant predictors of their judgments about how well students were likely to perform in several school subject areas.

While these studies show an influence of attitudes on language evaluation tasks, perceived accentedness and expectations of performance are meta-linguistic or non-linguistic judgments. These types of tasks can address whether beliefs and attitudes about speakers' social characteristics can affect judgments about language or other characteristics of speakers, but not whether beliefs about these characteristics can influence the *perception* of speech. However, social information has also been shown to influence more implicit judgments about speech sounds.

Plichta and Rakerd (2002) showed that listeners interpret the same auditory signal as a different phone when the speech in which the token is embedded gives them different clues about the dialect background of the speaker. Listeners heard the target vowel, about which they provided phone categorization judgments, in contexts that included other vowels that signaled the speaker's dialect. In the speech that provided the context for the target vowels, some listeners heard evidence of participation in the

Northern Cities Chain Shift. Plichta and Rakerd's study focused on the change in which /a/ is raised and fronted to sound like /æ/. Speakers in Lower Michigan, who live in close proximity to Detroit, participate in this change, but speakers in Upper Michigan do not. Because this vowel shift can make some words sound like other real words in other dialects, not knowing what dialect somebody is speaking can create ambiguity. Target words contained synthesized vowels that were on a continuum between /a/ and /æ/ (the vowels in the words *socks* and *sacks*). Lower Michigan listeners who heard speech showing evidence of participation in this change (produced by a Lower Michigan speaker) judged the same ambiguous sound clip to contain a different vowel than those who heard speech spoken by an Upper Michigan speaker.

Although no overt information was given about speakers' dialect background, listeners were able to make inferences based on the other vowels in the sample. It is plausible that listeners categorized speakers as having certain social characteristics (such as being from one geographical region or the other) based on the information contained in their vowels and then used this top-down information about speakers to make social judgments. Alternatively, however, listeners could have been using language-internal probabilities when making these judgments. This kind of effect could be caused by a normalization of the vowel space based on formant ratios, for example, rather than top-down inferences about different types of speakers. In this case, it would not necessarily be a case of making inferences about speech based on social characteristics – rather, it could be a case of making inferences about speech based on other aspects of speech.

Some documented influences of social information on language comprehension, however, strongly suggest that social factors themselves, and not language-internal factors, affect language comprehension. Evidence from event-related potentials (ERPs) indicates that listeners are influenced by socio-pragmatic information from the speech signal when they comprehend language (Van Berkum, Van den Brink, Tesink, Kos, and Hagoort, 2008). Hearing a phrase like *I just quit smoking* in a child's voice or *I'm pregnant* in a man's voice induces an ERP response indicating surprise. These phrases are pragmatically incongruous in combination with the social information implicit in the voices; the combination of the semantics of the phrases with the social characteristics of the speakers and the listener's world knowledge generates the surprise. The ERP response the authors observed is very similar in magnitude, direction and timing with one usually associated with semantic anomalies (known as the N400, a negativity appearing approximately 400 ms after the onset of the anomalous material), which can be induced by sentences such as *She spread her warm bread with* ***socks*** (Kutas and Hillyard, 1980). This suggests that listeners are considering the plausibility of an utterance given inferences they have made about the speaker during sentence comprehension on the same time scale that they use information about the semantics of other words in the sentence. Although the information about speakers' social characteristics comes from their voices, this effect is on the level of message, rather than the level of phonetics. Because of this, it is hard to imagine a normalization mechanism that involves only language-internal factors and not the top-down influence of world knowledge that could accomplish this feat. However, the effect found in this study (the N400) is not specific to language, and while the social information in the voices is certainly influencing the listener's response to the utterances, this experiment does not directly show that listeners are making inferences about speech based on this information. The results are equally consistent with the idea that listeners are making inferences about social characteristics of speakers based on the linguistic information, and it is the conflict between these inferred speaker qualities and those that they inferred based on the speaker's voice that yields the 'surprised' N400 response.

Using a more linguistic measure, like the phone categorization paradigm used in the Plichta and Preston study above, does not always eliminate the problem of determining the direction of inference. In another investigation of social information from speakers' voices influencing vowel perception, Drager (2006) found that there is a correlation between listener perceptions of /æ/-raising (a change in progress in New Zealand associated with younger speakers) and listener perceptions of speaker age. She played listeners a continuum of vowels from *bad* to *bed*, asked them to judge which vowel category the stimuli belonged to, and afterwards asked them to judge the speaker's age on the basis of the same tokens. Listeners judged higher tokens to be tokens of *bad* for the speaker who was, on average, rated to be younger. It is hard to establish the direction of the causal arrow in this scenario, because either the fact that the speaker sounded older could have influenced listeners' vowel perceptions, or the fact that listeners perceived more vowel raising could have influenced their estimates of the speaker's age, or they could both have been influenced by a third factor in the speaker's voice. A non-significant trend noted in the paper suggested that the correlation was not a coincidence – participants who had reversed judgments on age also had reversed perceptions of vowel raising, such that the speaker who was perceived as younger was always perceived to have more raising. While it was not statistically reliable, this trend suggests that the relationship observed between social characteristics and speech was not a coincidence. However, because neither factor was explicitly manipulated, these results can only establish a correlation between the two factors; as such, this study cannot distinguish between listeners making inferences about social characteristics (i.e., age) based on linguistic factors (possibly vowel raising, or at least some features of the spoken stimuli), and the opposite.

Providing social information explicitly, rather than through the linguistic signal, can make the direction of inference clearer. Niedzielski (1996, 1999, 2001) showed that listeners' beliefs about the geographical origin of a speaker affected their perception of the speaker's vowels. She played samples of speech to American participants from the Detroit area and told half of the listeners that the speaker was from Detroit and the other half that the speaker was from Canada. Speakers from both Detroit and Canada sometimes show evidence of a dialect feature known as Canadian Raising, in which the nuclei of the low diphthongs [ay] and [aw] are raised to the more central [ʌ]. While Americans from the Great Lakes region sometimes use this raised diphthong, it is traditionally associated with Canadians, and represents a departure from most American dialects. When listeners believed that the speaker they were hearing was from Detroit, they perceived the speaker's vowels to be more standard, in American terms: they assigned vowels to a low category, indicating that they perceived less Canadian raising. When they believed the speaker of the same token was Canadian, they assigned the same vowels to a higher category, demonstrating that they perceived more Canadian raising. Listeners' perceptions of phones were affected by their preconceptions about the way the speaker would talk. These results suggest that listeners' stereotypes about Canadians affected the way they categorized phones. Although providing the social information explicitly clarifies the direction of the inferences involved, because this information was explicitly given, it may have been subject to task demands – if the information was given, participants probably believed it was relevant to the task and may have felt pressure to use it. Thus, this task does not give us information about how listeners use social information that they happen to encounter while understanding speakers' utterances, when there are no demands to use the information.

In an extension of Niedzieski's findings, Hay, Nolan and Drager (2006) showed that the effect of labeling speakers' dialect region on vowel perception may not be based on conscious, top-down strategies. Listeners from New Zealand participated in a phone identification task where they heard the vowel in the word *fish*, which is pronounced more like *fush* in New Zealand and more like *feesh* in Australia. Participants reported hearing raised /ɪ/, which is more consistent with the production of an Australian, when their answer sheet had the word "Australian" at the top of it, but they reported hearing a centralized /ɪ/, which is more consistent with the production of a New Zealander, when their answer sheet had the word "New Zealander" at the top of it. The simplest potential mechanism behind this effect might be that listeners changed their beliefs about the nationality of the speaker, which would be consistent with Niedzielski's (1999) findings about Canadian raising, in which presumably participants changed their beliefs about the nationality of the speaker in response to being told what the speaker's nationality was. However, Hay et al. report that nearly all participants believed that they were listening to a New Zealander speak, even if they were in the Australian labeling condition. This suggests that social information can play a role below the level of conscious inference, since the effect does not appear to be rooted in an actual change in participants' beliefs about the nationality of the speaker, at least to the extent that they could report those beliefs. However, this finding is subject to the same limitation that Niedzielski's were – the presentation of the social cue explicitly by the experimenters may have created a task demand to use the information that does not exist in natural social language understanding situations.

In a follow-up to the 2006 study, Hay and colleagues elicited a similar effect just by having plush toys in the shape of kiwis (in the New Zealand condition) or kangaroos and koalas (in the Australia condition) in the room with participants (Hay and Drager, Under Review). The experimenter pretended to be surprised when encountering the plush toys, and presented them as unrelated to the experiment, while ensuring that all participants noticed their presence. In this case, task demands are an unlikely explanation for the effect. The authors interpret the persistence of the effect in this condition as indicating that activating the region of Australia or New Zealand through the plush toys causes previously heard exemplars of the speech of speakers from the relevant region to become active through spreading activation. Exemplar theory has been invoked to account for several results in speech perception showing the influence of the individual speaker or groups of speakers on how listeners categorize speech sounds; the current study is discussed in terms of this account in Chapter 6. This effect may ultimately be caused by the same mechanism, at some level, as inferences from speaker characteristics to speech; however, this study clearly does not support the conclusion that listeners make inferences about speech based on speaker characteristics, because the 'social information' they encounter is explicitly not about the speaker at all. In eliminating task demands to use the presented social information, the authors have also eliminated any sensible opportunity for listeners to make inferences. However, the existence of this non-inference-based effect suggests that the mechanism by which social information about speakers influences language comprehension would not have to be strategic, and could be entirely automatic.

Eliminating task demands to use social information does not, however, require eliminating the conditions necessary for sensible inference. Strand (1999, 2000) conducted a set of experiments that also avoids task demands to use social information, and showed that listeners are sensitive to both visually presented information and information in the speech signal about speaker gender in phone discrimination tasks. Gender differences (which may be due to differences in vocal tract size) correlate with differences in the production of some phones, such as fricatives [s] and [ʃ], in which productions by adult females tend to have higher frequency spectra than productions by adult males (Schwartz, 1968; Jongman, Wayland, and Wong, 2000). Investigating the perception of these phones, Strand found that seeing a picture of a man or a woman affects how people categorize ambiguous stimuli between sibilants /s/ and /ʃ/: when listeners are presented with a picture of a female speaker, they place the boundary between /s/ and /ʃ/ at a higher frequency than when they are presented with a picture of a male speaker. In addition, the gender stereotypicality of the voices affected listeners' phone discrimination – voices that sounded more stereotypically male elicited lower phone boundaries than less stereotypically male voices, and more stereotypically female voices elicited higher phone boundaries than less stereotypically female voices. Johnson, Strand, and D'Imperio (1999) showed that gender stereotypicality (of voices and photos) also affected the perception of the back vowels [ʊ] and [ʌ] (as in *hood* and *hud*) in a similar experiment, a variable that also correlates with gender in production.

Using pictures to present social information has the benefit that participants may not be aware of which social characteristic(s) are intended to be relevant to the task, reducing demand characteristics. In Strand's studies, participants may not have been aware that gender was relevant to the experiments in any way. However, the effects reported in these studies did depend on gender, which is a variable that has both social and physiological components that might influence phone production. The linguistic variables investigated in these studies are associated with gender partly through vocal tract characteristics, so it's unclear to what extent the effect relates to sociolinguistic knowledge and to what extent it is part of the speaker normalization process, which may use gender as a heuristic for vocal tract size (although inaccurately sometimes). Both social and physiological factors appear to be involved in gender differences in the production of these variables (Fox and Nissen, 2005). This leaves open the question of whether purely social information about speakers has the same influence on phone perception that Strand and colleagues observed from gender information. That is, can social information influence speech processing, per se, independently of the cues social information can provide about vocal tract physiology?

Hay, Warren, and Drager (2006) found influences of cues to age and social class on the perception of vowels involved in a merger-in-progress. They investigated the NEAR-SQUARE merger in New Zealand English, which is more advanced in younger speakers and in working class speakers. They showed listeners pictures of old and young speakers, and speakers in a working class and upper class guise, and discovered that for speakers who maintained a distinction themselves, seeing an older speaker made them more sensitive to the difference, compared to seeing a younger speaker. The social class manipulation had a more complex effect, interacting with both the amount of distinction produced by each speaker and the amount of distinction produced by the listeners. Like Strand's experiments, this study shows an influence of social information about speakers on the perception of phones associated with different social groups in production; however, in this experiment, there are no physiological relationships between the linguistic and social variables involved, making the effect purely social in nature.

Unfortunately, it is very difficult to determine what inferences listeners might have been making based on speaker age by measuring their sensitivity to the difference between these partly merged vowels. If listeners expected more difference between vowels for older speakers, they might have judged an equivalent distinction to be larger for older speakers than younger speakers, because they found it easier to perceive a difference when it was more expected. Alternatively, if listeners expected more difference between vowels for younger speakers, they might have judged an equivalent distinction to be more different for older speakers than for younger speakers, because they found an unexpected difference more salient. Because both these mechanisms are plausible, the fact that listeners perceived more distinction for older speakers than for younger speakers doesn't determine what expectations they had based on the speaker's age.

In addition, the fact that the difference in perception of the distinction occurred only in participants who maintained a distinction in production themselves raises the possibility that seeing older speakers made participants think of how they would speak to such an interlocutor, for example. For participants who maintain a distinction, this might cause them to simulate speech with a greater distinction, because this more conservative production is more formal. For participants who don't maintain a distinction, simulations of speech to any interlocutor will result in representations of fully merged vowels, since that is the only system to which they have access. In this scenario, participants' performance would be influenced by their own plans for speaking, and not by their inferences about the pictured person's speech.

Although it is difficult to conclude for certain that listeners are making inferences about speech from social characteristics of speakers from any one of these studies, the fact that several different approaches have yielded converging evidence suggests that this conclusion may indeed be warranted. However, a common theme across investigations of purely social information influencing speech perception *per se* is that the tasks used involve phone categorization (generally in a two-alternative forced choice paradigm). In these paradigms, listeners' attention is drawn to the specific set of two alternatives that are of interest to the researcher. In addition, although some of these experiments do not provide explicit task demands to use the information of interest, there may be general task demands to use sources of information that listeners do not normally use, due to the absence of cues such as semantic context, which may normally do most of the work in disambiguating ambiguous structures. Because explicit phone categorization is not a part of normal language comprehension, these studies leave open the question of whether people use information about speakers when they're not making a two-way judgment about language, but actually trying to understand a speaker in real time, in the presence of more contextual information. Is social information one of the clues listeners use when figuring out the puzzle of spontaneous speech? The first step toward answering this question is finding out what listeners know about the relationships between social characteristics of speakers and the speech they produce.

Chapter 3

The Variable: *t/d* deletion

This chapter provides background information about the sociolinguistic variable t/d deletion, which is the test case for the questions I am addressing in this dissertation. In the beginning of this chapter, I define the variable and discuss why it is a good choice for the current studies. In section 3.1, I provide a brief survey of the factors (both linguistic and social) that condition production of *t/d* deletion. Section 3.2 discusses the small body of literature relevant to the perception of the variable.

While this dissertation ultimately concerns the accumulation and use of social knowledge in general, these questions are most easily approached one variable at a time. The ideal variable for a study of sociolinguistic comprehension is one that has been well-studied from the point of view of production. Storing information about correlations in production is a plausible way in which sociolinguistic knowledge might be acquired. Because of this likely relationship between sociolinguistically conditioned production and sociolinguistically conditioned comprehension, the facts about the social conditioning of the production of a variable are a good source of hypotheses about what listeners might know about that variable. Also, "external" sociolinguistic factors that condition variable linguistic production do so in conjunction with "internal" phonological and/or morphosyntactic factors, which also influence phonetic variation (Labov, 1994; 2001). A good understanding of both external and internal factors influencing variable production is the basis for making predictions about how listeners will make use of social and linguistic information to make predictions about spoken language.

The variable this dissertation focuses on is called *t/d* deletion, a specific case of consonant cluster reduction. It is a phonetic variable in English in which final coronal stops in consonant clusters may be deleted in some environments (*fas(t) car; ban(d) practice*). Final *t/d* deletion is defined as the absence of a pronounced oral stop segment corresponding to a final *t* or *d* in words (Gregory et al. 1999). The possible realizations of the final consonant vary along a continuum from an aspirated *t* with a strong release burst to a completely deleted *t,* which leaves few spectral remnants in the acoustic signal. Without this information, there is no acoustic cue to the final consonant in the phonemic representation of a word.

Although many discussions of *t/d* deletion consider only fully present and fully deleted variants, it is not the case that there are only two possible realizations of the final stop, present or absent. Speakers produce stops that vary continuously from fully realized and released, on the strong end of the scale, to fully deleted on the weak end of the scale, with unreleased *t*, glottalized *t*, glottal stop, flap, and other realizations in between (Podesva, 2003). While there is evidence that many different aspects of the phonetic realization of this variable may be socially meaningful (Podesva, 2006), the standard taxonomy of consonant cluster reduction distinguishes only between two variants: the deleted and non-deleted variants. The non-deleted variant is any realization that includes a closure for the consonant, whether it is released or not; the deleted variant is any realization that does not include a closure for the final consonant. However, even a realization with no apical closure often has other cues to the presence of an underlying /t/, such as the duration of previous segments, glottalization, or perturbations of frication noise or nasal formants in surrounding consonants. Listeners are very sensitive to subtle acoustic cues, even those that do not correspond precisely in time to the relevant segment (Nguyen and Hawkins, 1999), and the lack of closure may not prevent listeners from detecting the underlying /t/ in a 'deleted' token. Despite the fact that the range of possible /t/ realizations makes it challenging to stick to this standard two-way taxonomy, these two categories of consonant realization have been studied extensively from the perspective of production. Therefore, I consider a binary distinction in the experiments presented in this dissertation. Experimentally, this binary distinction is preferable to a more fine-grained set of consonant realizations because it allows for the largest possible difference in cues to social features of a speaker from speech.

In addition to allowing me to make the best use of the information about conditioning of *t/d* deletion in production, using a binary distinction had another advantage. It was necessary to have one of the conditions of this study involve an entirely deleted final stop, because it is only under this condition that a very convenient property of consonant cluster reduction manifests itself. When the final stop is completely absent, the deletion can sometimes cause ambiguity between two words. For example, the word *mast* produced without its final consonant becomes ambiguous with the word *mass*. As mentioned above, in natural cases of deletion there are often durational differences or other cues to an underlying /t/, but it is possible to create tokens that could be either *mass* or *mast* without any of these cues by having naïve readers produce tokens of *mass*, which never contain any cues to a /t/, because there is no underlying /t/ in this word. This situation provides a good opportunity to see the effects of contextual information on language understanding, because the resolution of this ambiguity is a situation in which listeners would be well-served by taking contextual information into account.

This variable also makes a good test case for the current hypothesis because it has been one of the most commonly studied variable phenomena in spoken English. Final stops in consonant clusters are common enough that *t/d* deletion can be studied even using small collections of speech, and realizations of final *t/d* are variable in most groups and in most situations. Because of these qualities, an investigation of *t/d* deletion has been included in a part of a large number of community studies of variation, and it has also been a case study for many aspects of variation in phonology (Bayley, 1991; Guy, 1980; 1991; Guy and Boberg, 1997; Guy and Boyd, 1990; Hudson, 1997; Labov et al 1968; Labov, 1975; 1989; Neu, 1980; Patrick, 1999; Reynolds, 1994; Roberts, 1995; Santa Ana, 1996). Thus, information about the conditioning factors, both internal and external, that influence the use of this variable is readily available and fairly complete, across many different populations in the U.S.

3.1 Conditioning of *t/d* Deletion

Consonant cluster reduction is conditioned by several aspects of the linguistic environment, both phonological and morphosyntactic (Fasold, 1972; Labov et al., 1968). Features of the segment before the stop can encourage or discourage deletion; if the segment preceding the stop is a fricative (as in *last night*), deletion is more likely than if this segment is a liquid or a nasal (as in *cold night*), and deletion is least likely if the preceding segment is another stop (as in *rapt lover*). Features of the segment following the stop also influence deletion. If the segment following the stop is a consonant (as in *fast car*), deletion is most likely; if the following segment is a vowel (as in *fast action*), deletion is less likely, and if the stop is at the end of an utterance (followed by a pause), deletion is least likely. Stronger (more perceptually salient) variants of *t/d* appear more frequently in a cluster after stops, next after fricatives, and least frequently after sonorant consonants (Podesva, 2003). Podesva suggests that this may be because clusters with other stops provide very few perceptual cues to factors like place of articulation and durational cues for voicing. Thus, the fewer cues provided by the phonological environment to the identity of the stop, the more likely a speaker is to produce a full realization of the stop itself.

The morphological status of the stop also affects the likelihood of consonant cluster reduction. Stops that are part of the same morpheme as the previous segment (like the *t* in *past resolution*) are more likely to be deleted than stops that form their own past tense morphemes in phonetically identical words (like the final stop in *passed resolution*). Stops that form irregular past-tense markers (like the *t* in *kept*) can pattern either with monomorphemes or with other past tense markers, depending on age, dialect region and even on the style speakers are using (Guy and Boyd, 1990; Lim and Guy, 2005). Morphologically complex words may have less deletion than monomorphemic words because in morphologically complex words, the final stop carries tense information (Guy 1980). Combined with the facts about phonological conditioning, this suggests that overall, stops are deleted more when they are less informative and deleted less when they are more informative.

While the linguistic factors conditioning *t/d* deletion do so in a structured and fairly consistent manner across speech communities, they do not account for all the variability in consonant realization. Consonant cluster reduction is also conditioned by many stylistic and social factors, and this conditioning has been studied extensively in a variety of communities. In the case of this variable, as with many reduction phenomena, one variant (the retained *t/d*) is considered more standard than the other (the deleted *t/d*). The retained variant is usually considered more standard not by coincidence, but because of language ideologies that assign higher status to forms that are more historically conservative. This perception may be due to the fact that change is often motivated by ease of articulation, leading people to perceive new, reduced forms as lazy and imprecise (Kroch, 1978). Eckert describes this ideological pattern as "the association of hyperarticulation with care and hypoarticulation with laziness" (2008, 12). Perhaps due to the prevalence of these ideologies, there are several patterns of social conditioning that commonly apply to reduction phenomena, and apply to some degree to *t/d* deletion.

First, gender influences rates of *t/d* deletion. Men delete final stops in consonant clusters more often than women do (Wolfram, 1969). This is an example of the general phenomenon in which women tend to produce more standard, less reduced forms of stable and morphosyntactic variables than men do, although this pattern does not apply to all women (Wolfram and Fasold, 1974, p.93). Several accounts of this have been proposed, most of which ultimately depend on the fact that in the cultures in which this pattern has been observed, women have less access to power and have lower social status than men do (Eckert, 1989b; Trudgill, 1972). Thus, their use of more standard linguistic forms has been described as a way of accruing *symbolic capital* (Bourdieu, 1984), which is a source of societally-granted status that is available even to those who do not have much economic or political capital (and, as in the case of women in the Western world, may be the type of capital they have the most access to).

In addition, age influences use of these variants, with younger people tending to delete more often than older people do (Guy and Boyd, 1990). This is also a common pattern for reduction phenomena, perhaps due to the fact that formality has different social implications for younger speakers. The reduced variant is generally more casual than the alternative (see below), and younger speakers use more casual forms. It is also possible that younger speakers use their vernacular more than older speakers do.

While it is not always the case that race or ethnicity influences reduction phenomena, in this case race is an important conditioning factor on consonant cluster reduction. In particular, African Americans delete final stops more often than Caucasian American speakers do (Rickford, 1999; Wolfram, 1969). Because it is used more by African American speakers than by Caucasian American speakers, and is a well-studied feature of Black speech, *t/d* deletion has sometimes been described as a feature of African American Vernacular English (commonly referred to as *AAVE*) (Fasold, 1972; Wolfram, 1969). However, an important finding of Wolfram's Detroit study, among others involving *t/d* deletion, is that upper middle class African Americans have very high rates of *t/d* deletion (nearly 80% in some phonological environments – see Figure 3.1), even if they use very few other dialect features of AAVE (Wolfram, 1969). This suggests that although *t/d* deletion may be a feature of AAVE, it is also a feature of non-vernacular Black Englishes.

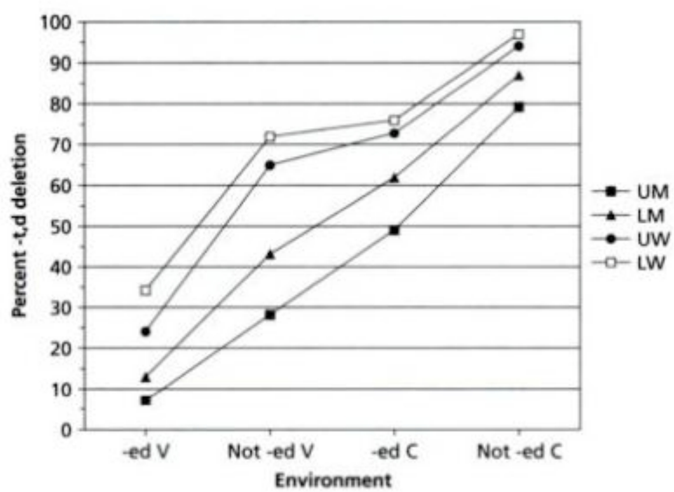

Figure 3.1. Among African American speakers in Detroit, Wolfram found that although there was social stratification in the use of t/d *deletion, the differences between phonological environments were larger than the differences among social classes (reproduced from Wolfram and Fasold, 1974:132).*

Because its use is so widespread among African American speakers of all social classes in Wolfram's sample, and also appears in all American varieties of English, it does not seem appropriate to consider *t/d* deletion an AAVE marker, but rather a feature of many varieties of American English that appears at a particularly high rate in the varieties used by African American speakers. However, several studies of African American communities in recent years (Wolfram, Hazen and Tamburro, 1997; Mallinson and Wolfram, 2002; Carpenter, 2005) suggest that consonant cluster reduction, especially in certain phonological environments such as prevocalically, remains an ethnolinguistic marker and is reliably more common among African Americans than their European American peers (see Figure 3.2).

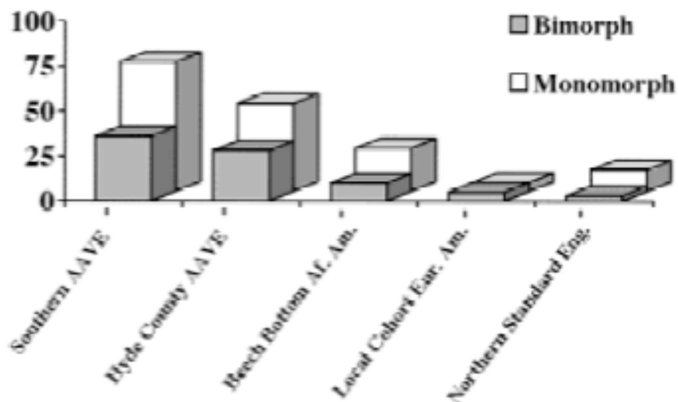

Figure 3.2. Rates of consonant cluster reduction were higher among African American populations than European American populations, especially in bimorphemic words, in recent studies of communities in North Carolina (reproduced from Mallinson and Wolfram (2002)).

The distinction between being a feature associated with African Americans and being an AAVE marker is important for the experiments in this dissertation, because the cues to race that were provided to participants were photographs from a University of

Pennsylvania ID photo database. The Black males whose pictures were chosen for the experiments may not have consistently appeared to be speakers of any particular African American English variety, but because high rates of *t/d* deletion are common to varieties spoken by African Americans of all classes and education levels, participants should not need to ascribe a particular dialect to the purported speakers in order to make the inference that their rates of *t/d* deletion would be higher than those of their White counterparts.

Another advantage of using *t/d* deletion as a test case is that while many sociolinguistic variables differ in meaning from one speech community to another, there are a few sociolinguistic variables that have some consistency in meaning across the U.S. Consonant cluster reduction has several linguistic and social conditioning factors that are fairly consistent from community to community. For example, though regional differences have been found between AAVE-speaking communities in the use of other phonological variables, such as /r/ production (Hinton and Pollock, 2000), *t/d* deletion is a feature that has been consistently identified as associated with African Americans in community studies, and reduction phenomena were among the few variable phenomena that were a consistent feature of African American Englishes in a large-scale spoken corpus study of dialect differences (Schwartz et al., 2007). Thus, probing the intuitions and knowledge of a diverse population like the Stanford community, many of whom were raised in other regions of the U.S., should still result in a coherent picture of the listener's knowledge of *t/d* deletion. However, the fact that these patterns in *t/d* deletion are specific to the U.S. as a speech community requires that participants must have grown up speaking English in America – speakers of other global varieties of English may have different associations with *t/d* deletion, but are unlikely to have the same knowledge of this variable's association with African American speech that Americans might have, upon which the studies in this dissertation are based.

In addition to its social conditioning, *t/d* deletion is also a stylistic variable, which means that its use is conditioned by the circumstances of the speech it appears in. The deleted variant appears more often in casual, more vernacular speech than in careful speech (Baugh, 1979; Guy, 1980; Labov, 1972). Lim and Guy (2005) found that in addition to showing differences in overall rates of *t/d* deletion, different styles also sometimes show differences in the ranking of morphophonological constraints. According to this study of speakers of Singaporean English, the constraints on *t/d* deletion operate differently in different styles. Some of the differences they report are in morphosyntactic constraints, and others are due to style-related changes in segmental phonology. These results suggest that restricting the experimental items to a single style may yield more consistent results.

The conditioning factors of *t/d* deletion make this variable a good choice for the current investigation, for several reasons. First, both variants of *t/d* deletion are used by people of all ages and races/ethnicities, and both genders, but with different frequencies. As such, a deleted final stop is less expected in certain social contexts, but it is not entirely unexpected or inappropriate in almost any circumstance. Perhaps because it is a phonological variable that is not exclusively associated with any group, people do not frequently cite *t/d* deletion when describing differences in the way people talk, which limits the effect that explicit stereotypes about language will have on the study.

The fact that race is a very robust conditioning factor of *t/d* deletion is also an advantage for the current purposes. It is possible for listeners to gather information about the race of a speaker from visual information, such as a video or a picture (or, in more natural conversation, the appearance of their interlocutor). In order to investigate the effect of speaker race on listener expectations about *t/d* deletion, it is possible to manipulate the purported race of a speaker by showing pictures of different speakers with the same sound clips, as in Rubin and Smith (1990).

For these reasons, it is the relationship between race and consonant cluster reduction that I examine in this dissertation. The speakers of the target sentences in the current experiments were selected to be young males, who have higher rates of *t/d* deletion than their older or female counterparts (Wolfram, 1969). Because the speech was produced and comprehended in a laboratory setting, the more likely and natural the deleted tokens sound, the more listeners are encouraged to process these as casual utterances with deletions.

Linguistic variants that are used more often by Black speakers are correlated not with a race but with a speech community that tends to have members of certain races. However, listeners do not have direct access to information about what speech community a speaker is a member of, and instead must make inferences based on observable characteristics of speakers that are associated with membership in a particular speech community. Current theories of race in America are based on the idea that racial categories are largely socially constructed, but may have some relationship to genetic or ethnic groupings (Omi and Winant, 1986). Although there may be few factors in the world that a priori determine the structure of racial categories, the social construction of these categories has led Americans to associate certain physical and social characteristics of people with one category or another (Hartigan, 2005). Thus, in the real world, listeners might use any number of features of a person to infer their race and thus the speech community to which they likely belong, including (but not limited to) their physical appearance, facial expressions, body posture, hairstyles, clothing, etc. In the experiments I present in this dissertation, participants had access to all of these features in the photographs of potential speakers they saw, but they did not have access to any more direct cues to speech community membership. I use the term *race* in this dissertation to describe the category structure that separates one group of potential speakers from another, because the cues I provided to participants in all six experiments were most consistent with race, rather than ethnicity or speech community. In addition, I use the terms *Black* and *White* to refer to the race of talkers in the experiments, and the terms *African American* and *European American* to refer to the ethnic groups in the world. The conclusions about the associations listeners form between types of people and kinds of linguistic variation are most likely formed on the basis of assumptions about speech community membership based on race.

3.2 Perception of coronal consonants

A few studies addressing the issue of how listeners perceive variation in the realization of coronal consonants provide further reasons for using *t/d* deletion in a study of sociolinguistic perception. Sumner and Samuel (2005) showed that words containing a phonetically reduced /t/ still prime semantically related targets. Words in which [t] was replaced by sounds that are regularly produced by speakers as realizations of /t/ such as glottalized [t] or glottal stop with no coronal articulation primed semantically related words, whereas words in which [t] was replaced by sounds that differed from it in some other feature such as manner of articulation (e.g. [s]) did not produce any semantic priming, suggesting that regular form reduction does not hinder semantic processing, although arbitrary phonetic changes of similar magnitude do. These findings indicate that regular phonetic reduction does not interfere with listeners' ability to recognize words, suggesting that it should be possible to investigate the sociolinguistic perception of *t/d* deletion using a paradigm relying on the meaning of words with deletions.

In addition, there is evidence from perception studies that the context of a reduced /t/ can influence whether listeners perceive a /t/ at all. Mitterer and Ernestus (2006) found that when presented with a lenited version of a /t/, Dutch listeners were more likely to report hearing a /t/ when it appeared in circumstances favorable to the lenition of /t/, such as after an /s/ (instead of an /n/); listeners were also more likely to report hearing a /t/ when the presence of a /t/ would create a real Dutch word and its absence would create a non-word, and they were less likely to report hearing a /t/

when its presence would create a non-word and its absence would create a real word. The fact that both phonological and lexical context can influence the perception of reduced /t/ in Dutch suggests that the perception of *t/d* deletion in English might also be influenced by non-linguistic contextual factors, such as social characteristics of speakers.

One study has addressed the social meaning of some variants of /t/. Campbell-Kibler (2005) conducted a pilot study comparing listeners' reactions to released vs. unreleased /t/, which may be related to the way listeners interpret *t/d* deletion, as it involves other variants on the same continuum. She found that released /t/ influenced how smart, casual, and relaxed listeners believed a speaker to be, but not how educated or wealthy. These results demonstrate that listeners are sensitive to cues as subtle as release when making subjective judgments about a speaker. However, because both variants she considered are non-deleted consonants, these findings do not shed direct light to listener perceptions of deleted vs. non-deleted *t/d*. While there may be a relationship between the meanings of unreleased *t* and deleted *t*, it is not possible to infer from this study the meanings listeners ascribe to the full spectrum of *t/d* realization, leaving this a largely unexplored area.

Evaluating listeners' knowledge of the relationship between race and *t/d* deletion will provide a window into how listeners store and use knowledge of sociolinguistic variation in their everyday language comprehension. Discovering the ways in which listeners learn and remember facts about specific sociolinguistic variables can help to fill in the missing link between socially conditioned variable production and the meaning that seems to accompany it. In the next chapter, I present two experiments investigating listeners' knowledge of relationships between *t/d* deletion and race, and how they might have acquired it.

Chapter 4

What do listeners know about *t/d* deletion?

The previous chapter described an abundance of information available from community studies on multiple aspects of the production of *t/d* deletion. These studies established that being young, being male and being Black are factors that make a speaker more likely to produce the deleted variant than a speaker who is older, female, or White (Fasold, 1972; Labov, 1966; Wolfram, 1969). Together, these studies have established that speakers tend to use the deleted variant more in informal speech situations and the non-deleted variant more in formal speech situations (Baugh, 1979; Guy, 1980; Labov, 1972), but, as discussed at the end of that chapter, very few studies have focused on how listeners interpret these variants.

The fact that many social factors correlate with *t/d* deletion in production implies that use of this variable may be socially meaningful. However, the fact that linguists can attend to and measure rates of *t/d* deletion and find correlations with social factors does not mean that listeners regularly do this. Campbell-Kibler (2005, 2007) provided some of the first investigations into the ways in which listeners find changes in single variables meaningful, and found interesting correspondences between factors governing production of a variable and the meanings it holds for speakers. However, in the case of *t/d* deletion, there is not yet evidence that the correlations between linguistic variation and social factors that have been discovered through production studies have any psychological reality for the listener.

In principle, listeners could attend to and store information about all the relationships that exist in the world between social characteristics and linguistic variation, resulting in a vast array of learned associations. This may in fact be the case; however, it is also possible that listeners are insensitive to much of this variation. Which variables they store information about could be constrained by a variety of factors. Perhaps listeners attend only to variation in cues that consistently provide information about social group membership. They might attend only to variation that is relevant to their social lives, or the present situation, in certain ways. Listeners might attend only to linguistic variables that they themselves use. The possible constraints on this attention are nearly limitless.

Because it is impossible to consider all sources of socially meaningful variation at once, a fruitful way to pursue the question of how sociolinguistic variation affects listeners' perceptions is to frame empirical questions in limited domains. The results of experimental investigations of individual variables can be combined to construct a larger understanding of how listeners map the landscape of sociolinguistic variation, and how this information is incorporated into the mechanisms of language understanding.

In this chapter, I present two experiments investigating whether listeners have implicit knowledge of the relationship between race and *t/d* deletion in American

English, and if so, how specific this knowledge is. Experiment 1a investigates whether people associate deleted final consonants with Black speakers more than they associate them with White speakers. Experiment 1b addresses whether this relationship might be due to a general belief that any nonstandard usage is more likely to be spoken by a Black speaker, or a more specific belief about *t/d* deletion. The experiments were conducted together, but are explained separately for the sake of clarity.

4.1 Experiment 1a

Experiment 1a examined whether listeners attribute *t/d* deletions differently to speakers of different races, based on the distribution of these variants in speech. If listeners store information about the distribution of deleted final consonants in the input in some way, they should associate deleted final stops more with Black speakers than with White speakers. This association would be consistent with the findings of the community studies from across the nation discussed in the previous chapter. In addition to being interesting on its own, the existence of this relationship in the minds of listeners is an important prerequisite for Experiments 2 and 3 - if listeners have knowledge about the distribution of the variants with respect to race, then it is possible to examine how this information is used in processing language and represented in listeners' minds. If listeners do not appear to store information about which types of speakers produce the different variants, then they have no way to link social characteristics of the current speaker to probabilities of hearing specific variants. If they cannot make this link, then it is likely they do not use this information to help them understand language.

Experiment 1a was a modified Matched Guise study (Lambert, Hodgson, Gardner, and Fillenbaum, 1960), in which minimal pairs of sentences are constructed to create guises that differ only in one dimension. The original Matched Guise Technique pitted entire language varieties against one another. In the current study, the guises were controlled and differed only in *t/d* deletion (e.g., the presence or absence of *t* or *d* word finally). While the technique as developed by Lambert and colleagues originally involved oral guises, written guises have also been used in this paradigm, when the variables or varieties in question permitted it (e.g., Bradac and Giles, 1988; Buchstaller, 2004; Kramer, 1974). Written guises have the advantage of allowing greater control over some orthogonal factors such as other phonological variables. Because it is possible to orthographically represent *t/d* deletion, Experiment 1 used written guises involving *t/d* deletion. Participants saw sentences that include a word ending in either a deleted or a non-deleted consonant, and judged which of two possible speakers is likely to have said the sentence. Each sentence was paired with one White and one Black potential speaker. If listeners have unconscious associations between deleted *t/d* tokens and African American speakers, they should more often rate sentences with deleted tokens likely to have come from African Americans than sentences with non-deleted tokens.

4.1.1 Methods

Participants One hundred eleven Stanford University undergraduates received course credit in an introductory psychology class for their participation in this study, which was distributed as one page in a larger packet of unrelated surveys. Participants were of both genders and a mixture of ethnicities, and most participants were between the ages of 18 and 22. While a majority of participants were American, there may have been some participants of other national backgrounds[3]. Because the class was not in the area of linguistics, most participants presumably had little or no linguistics training and were probably not aware of the status of *t/d* deletion as a sociolinguistic variable.

Materials Twenty-four sentences were constructed so that each included a word with a consonant cluster that could be subject to *t/d* deletion (e.g. *mast, least, wind*). The *t/d* consonants were primarily in phonological environments that promote consonant cluster reduction, as discussed in the previous chapter, such as following a homogeneously voiced fricative or a nasal and preceding a word beginning with a stop or a glide (e.g. *The mast probably lasted*…). Each sentence appeared in two versions, the standard version and the deleted (nonstandard) version. In the standard version, this word was presented with its normal orthography (e.g., *mast*; Figure 3.1, line A). In the deleted version, this word appeared with its final stop replaced by an apostrophe

---

[3] While participants who did not grow up in America might not have the relevant background to provide useful data in this experiment, their participation should not bias the results in any systematic way.

(e.g., *mas'*; Figure 3.1, line B), indicating a deleted final consonant (see the Appendix for a complete list of stimuli and fillers).

Participants received one of two questionnaires, A (standard) or B (nonstandard). Each questionnaire contained twelve target items and twelve fillers. Questionnaire A contained the standard versions of the target *t/d* deletion sentences (presented in normal orthography). Fillers for Questionnaire A were twelve similar sentences presented in normal orthography. Questionnaire B contained the deleted (nonstandard) versions of the target *t/d* deletion sentences (written with an apostrophe in place of the final consonant). Fillers for Questionnaire B all contained an unrelated nonstandard usage (see Appendix). Thus, all the sentences in Questionnaire B contained a departure from Standard Written English, expressed orthographically, whereas all sentences in Questionnaire A contained sentences in Standard Written English. Because the standard versions of the sentences appeared only in Questionnaire A and the deleted versions appeared only in Questionnaire B, the difference between standard and deleted versions was a between-subjects comparison. This prevented sentences containing *t/d* deletions from standing out as especially non-standard to any given participant, and prevented participants from being able to compare deleted vs. non-deleted versions of similar sentences.

Underneath each sentence in Questionnaire B, participants saw a 'translation' of the nonstandardism to ensure that all participants interpreted the stimuli as realizations of the same words. In the case of *t/d* deleted items, this consisted of the same word written in standard orthography (i.e., the word *mast* was written underneath the form *mas'*). Items were counterbalanced across versions of each questionnaire so that although each participant saw only twelve target items, all twenty-four target items were seen in both standard and deleted versions by some participants.

Above each sentence on both questionnaires were two pictures, which participants were supposed to choose between to select the more likely speaker of the sentence. Pictures of potential speakers were taken from a database of University of Pennsylvania ID photos (Killgore, Casasanto, Maldjian, and Detre, 2000), and included the shoulders and head of college-aged individuals, on a white or neutral background (Fig. 4.1). Four Black and four White male individuals were used for both target and filler items, with a variety of pairings of Black and White individuals. The photographs were restricted to males because males tend to have higher rates of *t/d* deletion overall than females do (as discussed in Chapter 3). Using photographs of males for the potential speakers made the deleted versions of the sentences more plausible, and also eliminated the problem of cross-gender pairings in which gender might play as much of a role (if not more of one) as race in determining participant responses. Each sentence had one White and one Black individual pictured above it. These pictures obviously may have differed from one another in other respects than race; however, race was the only dimension on which the photos systematically varied.

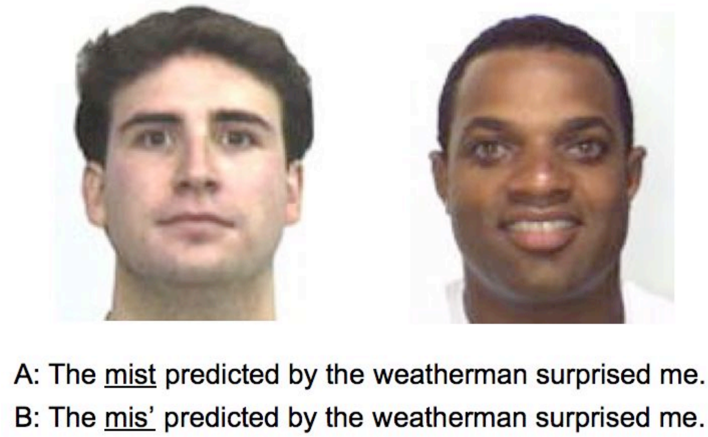

Figure 4.1. In Experiment 1, pairs of photographs were presented with either the standard (A) version or the deleted (B) version of the target sentences below them.

Procedure Each participant was randomly assigned to receive one of the two questionnaires. In both questionnaires, their task was to circle the picture of the person they thought was more likely to have said the sentence in question. The instructions for each questionnaire varied, as below:

Questionnaire A instructions:

> Below each pair of pictures is a sentence. For each pair, please circle the person you think is more likely to have said the sentence.

Questionnaire B instructions:

> The sentences below are transcribed from natural speech. Some of the words speakers used are non-standard. 'Translations' for these words are given in parentheses. Try to imagine how they might have sounded in your mind's ear. For each sentence, decide which person pictured above it is more likely to have said it.

58 participants received Questionnaire A; the remaining 53 participants received Questionnaire B. Using written stimuli allowed this experiment to address the influence of *t/d* deletion without the influence of auditory cues to race. Because participants' association between *t/d* deletion and race might not be conscious, participants were not made aware that either consonant cluster reduction or race was of interest. Other non-standardisms were present in the fillers to mask the target variable. However, due to the fact that all pairs of faces contained one White and one Black face, participants probably realized that race was of interest in the experiment.

4.1.2 Results and Discussion

Participants attributed 60% of the 'deleted' sentences, represented with apostrophes, to the pictures of Black students pictured, and the remaining 40% to the pictures of White students. By contrast, they attributed only 42% of the non-deleted sentences, with normal orthography, to the pictures of Black students, attributing the remaining 58% to those of the White students. This pairwise difference between proportions was significant ($t(1,109)=4.86$, $p<.001$) (Figure 4.2, below), indicating that participants do associate *t/d* deletion more with Black speakers than with White speakers.

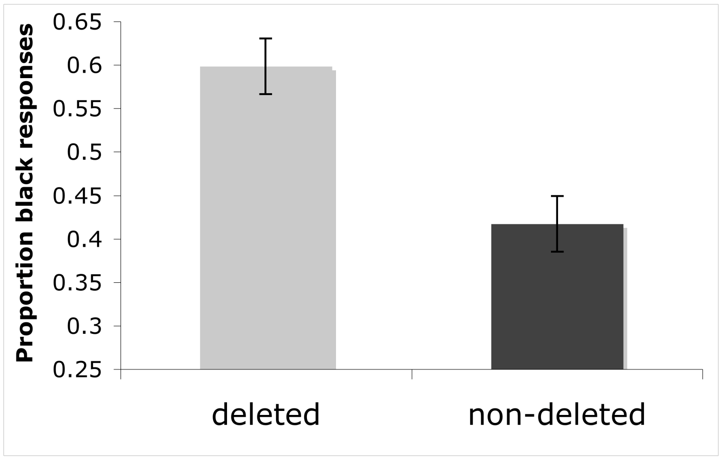

Figure 4.2. Participants in Experiment 1a selected the African American from the pair of pictures 60% of the time when they saw the deleted versions of the sentences, but selected the African American only 42% of the time when they saw the non-deleted (standard) versions of the sentences. Error bars represent standard errors of the means.

While the absolute frequency with which participants associated *t/d* deletion with Black and White speakers may be specific to the types of sentences the deletions occurred in, the fact that participants showed a similar association between *t/d* deletion and race to the association that occurs in production strongly suggests that they internalize probabilities related to social factors during language comprehension. However, it is possible that this association could have been learned without tracking probabilities; Experiment 1b investigated whether this effect could have been due to inferences based on stereotypes.

4.2 Experiment 1b

How specific is listeners' knowledge of links between language and race? Experiment 1a demonstrated that listeners associate speaker race with likelihood of producing deleted tokens. However, this fact alone does not necessitate that listeners have attended to variation in the world and stored information about it. As discussed in Chapter 2, demonstrating that listeners perceive an association between linguistic features and social characteristics does not by itself demonstrate that they have been storing information about correlations in their experience.

People have many beliefs about what kinds of people are likely to do different things, but these beliefs are not always based on experience with people actually doing things. The term *stereotype* refers to beliefs of this nature that can be based not just on experience but also on things like hearsay and media representations rather than direct experience with people. Because it's possible to hold beliefs about how people of different ethnicities, for example, will act without actually having observed them acting in this manner, evidence of a belief doesn't constitute evidence that people developed this belief through direct experience. The participants in Experiment 1 presumably had a variety of language backgrounds and experiences, and in this experiment they were not asked about their experience with African American speech. Thus, the results of Experiment 1a do not determine whether listeners believe that African Americans produce more *t/d* deletion, per se, or that African Americans produce more non-standardisms, in general. That is, these results do not rule out learning from stereotypes of African American speech, and do not argue in favor of direct experience with African American speech as the source of the beliefs identified in that experiment.

Could stereotyping, rather than direct experience with people of different ethnicities, be responsible for the results of Experiment 1a? In principle, this is plausible. Because the deleted variant is less standard than the non-deleted variant, participants could have associated *t/d* deletion with African American speakers simply because they associate less standard usages with African Americans. This would not require any experience on the part of the listener with the specific variable, only the knowledge that a) reductions and deletions are nonstandard, and b) African Americans are more likely to produce nonstandard language.

Experiment 1b investigated whether participants' beliefs about the relationship between *t/d* deletion and race are based on general stereotypes or linguistic experience. If participants associate less standard usages with African Americans based only on stereotyping, then they should show the same preference for selecting African American speakers with any other nonstandardism that they showed for *t/d* deletion, even if the other nonstandardism is not associated with African Americans in production. However, if they associate *t/d* deletion more with African Americans than European Americans based on experience with the variable itself, then they should show a greater preference for selecting African American speakers with *t/d* deletion than with other nonstandardisms that are not associated with African Americans in production, because African Americans actually produce *t/d* deletion more than European Americans do.

4.2.1 Methods

Participants 53 Stanford University undergraduates received course credit for their participation in this study, which was administered in a packet of unrelated surveys. Participants were of both genders and a mixture of ethnicities. A majority of participants were born and raised in the United States, but some may have grown up in other countries. These participants were a subset of those who participated in Experiment 1.

Materials Materials were identical to those used in Experiment 1a, Questionnaire B. While nonstandardisms served as fillers in Experiment 1a, in Experiment 1b, they were experimental items. Participants saw twelve sentences with an apostrophe representing a *t/d* deletion (the deleted version of the target sentences), and twelve sentences containing another nonstandardism that is not associated with African Americans. Four types of nonstandardisms were used, with three items of each type to produce the twelve items. These other types of nonstandardisms also appeared more than once so that *t/d* deletion was not unique in appearing multiple times in the questionnaire.

One of the four types of other nonstandardisms was another phonetic variant, the extreme diphthong in the New York pronunciation of words like *coffee* and *dog* ([kɔᵊfi]; [dɔᵊg]); these vowels were represented orthographically as *cawfee* and *dawg*. Like the *t/d* deletion items, these were ‘translated’ below the sentences by the standard orthography of the words, to ensure that the fillers were as similar as possible to the target items. Two of the other nonstandardisms were morphosyntactic. The first was the "needs washed" construction, as in *The car parked near my house needs washed*. These were 'translated' below the sentences as *needs to be washed*. The second morphosyntactic variable was double modals, as in *My twin brother might could pass for me at school*. This was 'translated' as *might be able to*. The fourth nonstandardism was lexical; the phrase *youse guys* appeared in the sentences, translated as *you guys*. Importantly, none of these variables is used more by African Americans than by European Americans (Labov, 1966; Murray, Frazer, and Simon, 1996; Smitherman, 1986).

The term *nonstandard* is used here as an umbrella term, encompassing different types of usages, for two reasons. First, the four 'other nonstandard phenomena' differ in the respects in which they depart from an imagined national standard – some are better described as vernacular, some are better described as regionalisms, and others as both. Second, the important way in which these usages form a natural class with *t/d* deletion is that they are not 'standard' in a sense that encompasses both of these qualities and others. While it is hard to know what constitutes the imagined national standard of our participants, all the usages included in Questionnaire B are less statusful than most of their alternatives, and this is all they must have in common with one another to be useful in Experiment 2.

The other non-standard usages were selected because they are not associated with African Americans in production, and this difference between the other nonstandardisms and *t/d* deletion is the intended one. However, the set of other nonstandardisms also differed from *t/d* deletion in other ways, perhaps most notably that it included some lexical and morphosyntactic features. Because not all nonstandard features are equally marked, it could be that these features were not equivalent to *t/d* deletion in how noticeable they were to participants. If *t/d* deletion was more noticeable to participants than the other types of nonstandard usages, the sentences containing this feature might seem more non-standard overall. If so, *t/d* deletion could be associated more strongly with African American speakers due to exactly the stereotyping effect that the results of Experiment 1b seem to argue against. However, language attitudes surveys have shown that non-standard lexical and/or morphosyntactic features tend to be more sharply stratified among speakers and more noticeable to listeners of a dialect than non-standard phonetic features are (Lippi-Green, 1997; Wolfram, 1969b). This pattern indicates that the difference in how marked or noticeable the other non-standard features were compared to *t/d* deletion would be a source of Type II error, working against the hypothesis[4].

Procedure The procedure and instructions for Questionnaire B, which contained the items used in Experiment 1b, were described in the Procedure section of Experiment 1a; participants read the sentences (and their 'translations') and judged which of the two people pictured above the sentence was more likely to be the speaker of the sentence.

---

[4] Type II error (or a 'false negative') is an error in which data do not seem to support rejecting the null hypothesis although in fact the alternative hypothesis is the true state of the world.

4.2.2 Results and Discussion

The sentences containing apostrophes representing t/d deletions were attributed to the Black speaker pictured above the sentence 60% of the time, whereas the other nonstandardisms were attributed to African Americans only 51% of the time. This pairwise difference between proportions was significant ($t(1,52)=1.97$, $p=.03$) (see Figure 4.3).

The attribution of the other nonstandardisms to African Americans was not significantly different from chance (50%), but the absolute percentage of participants attributing the sentences to African Americans may not be interpretable because of the small number of items and the large degree of variability between them. For example, one item that was intended to represent a New York City vowel (*dawg*) was associated strongly with African Americans by participants, presumably because this spelling has other interpretations besides the intended one. The *Urban Dictionary* (2008) defines this term as "Slang for 'my close acquaintance of an African-American ethnic background,'" suggesting that for many speakers, this term is explicitly associated with African Americans. While this is also a source of Type II error relative to the main hypothesis (because it would serve only to mask the difference between *t/d* deletion and other non-standard usages), it does inflate the absolute percentage of these other non-standard usages attributed to African Americans. However, there are at least two possible explanations for this pattern, if it merits interpretation. First, participants may not have associated these nonstandardisms more with one race than another because they believe that Whites and Blacks use them equally often. Alternatively, participants may have had no associations between these variables and race because they had little or no experience with these variables, and simply chose randomly.

Whether or not their responses to the other nonstandardisms were random, the difference between other nonstandardisms and *t/d* deletion indicates that participants' responses to the *t/d* deletion sentences in Experiment 1 was not simply a reflection of a general belief that nonstandardisms are more likely to be said by African Americans. Rather, it must have been based on some experience causing them to associate this variable itself with African Americans. This experience could have been in person or through media representations of African Americans using *t/d* deletion, but it must have been specific to this variable, rather than about nonstandard usages in general.

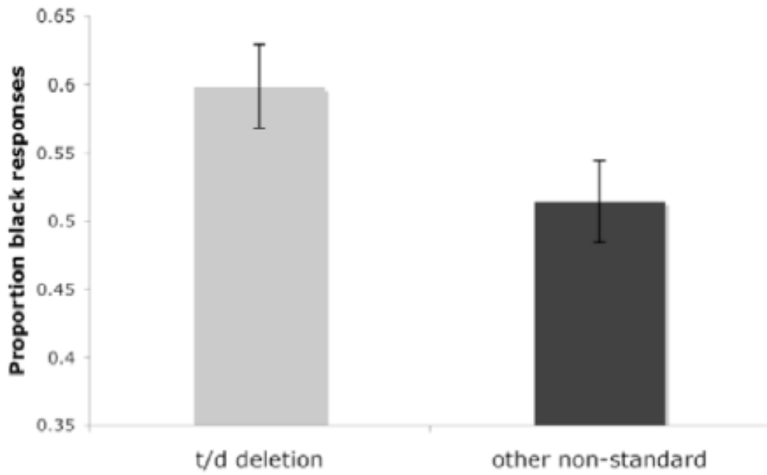

Figure 4.3. Participants in Experiment 1b selected the picture of the Black speaker 60% of the time when the sentence they saw contained a t/d deletion, but only 51% of the time when the sentence contained a different nonstandardism. Error bars represent standard errors of the means.

4.3 Summary

Experiment 1a demonstrated that listeners associate *t/d* deletion more strongly with Black speakers than with White speakers, which is consistent with the facts about production of this variable. Results of Experiment 1b indicate that although stereotyping may play a role in forming this association, listeners do distinguish between *t/d* deletion and other non-standard variants that they have likely never heard an African American produce. This suggests an important role for experience with linguistic material in forming the associations listeners have between social characteristics and linguistic variants. While these experiments confirm that listeners have access to statistical relationships between *t/d* deletion and race and suggest that experience with linguistic forms is involved in accumulating knowledge of these relationships, they leave an important question unanswered: do listeners use this knowledge when understanding language?

Previous studies (discussed in Chapter 3) have established that listeners can use both social information presented visually or in explicit instructions to participants, and social information that they must gather from an auditory signal to inform their expectations about how speakers will produce different phones. Thus, there is reason to expect that listeners should make use of social information to resolve naturally occurring ambiguity in on-line language comprehension.

There is also a more general reason to expect that social information might be useful in on-line comprehension. Chater et al. (2006) suggest that human language comprehension may be an optimized process:

> …it seems increasingly plausible that human cognition may be explicable in rational probabilistic terms and that, in core domains, human cognition approaches an optimal level of performance. (Chater, Tenenbaum, and Yuille, 2006, p. 289)

This view suggests that listeners should make use of any statistical regularity in their linguistic input to understand speech. The sociolinguistic relationships that listeners learn for the purpose of making social inferences about their interlocutors may be available to be co-opted for other purposes, as well. Thus, if listeners are using an optimal strategy to understand language, they should exploit social information as a source of predictability to solve the problem of speech perception and language comprehension. This prediction of an optimal view of language processing is empirically testable: if people use probabilistic information from the social context in the same way that they use information from other aspects of context in language processing, to optimize performance, then it should be possible to observe effects of social information on-line in language-processing tasks like ambiguity resolution. The next chapter presents two experiments testing this prediction in the domain of *t/d* deletion.

Chapter 5

Do listeners use social information during online language processing?

Experiments 1a and 1b showed that listeners have at least implicit knowledge that speaker race is correlated with *t/d* deletion. However, these experiments do not tell us whether listeners make use of this knowledge when they are engaged in the process of language comprehension. Experiment 2 investigated whether listeners use the associations between race and *t/d* deletion observed in Experiment 1 automatically in language processing tasks.

5.1 Experiment 2

Experiment 2 investigated this question using pairs of words that can be temporarily ambiguous when spoken aloud. The source of the ambiguity between the two words in each pair is the deletion of the final *t* or *d* segment of a word (*t/d* deletion), such that, for example, the word *mast* becomes confusable with the word *mass*. Each pair contains one word (the *t*-word) that has a final coronal stop (*t* or *d*) and one word (the non-*t*-word) that contains the same string of phones as the first word, but without the final coronal stop. When the *t*-word is pronounced, the final coronal stop may be deleted, yielding a phone string that is identical to that of the standard pronunciation of the non-*t*-word.

In Experiment 2, listeners were exposed to these ambiguous words and had to come to an interpretation of them. If listeners use their knowledge of sociolinguistic variation when they understand sentences, then they should be more likely to believe that a consonant cluster reduction has taken place when they believe the speaker is Black than when they believe he is White. Thus, they should interpret the word as having an underlying *t* more often when they believe the speaker is Black than when they believe he is White, because reaching this interpretation involves inferring a deleted stop. By contrast, they should interpret the word as not having an underlying *t* more often when they believe the speaker is White than when they believe he is Black, because reaching this interpretation involves rejecting the alternative with a deleted stop.

5.1.1 Methods

Participants 40 native U.S. English speakers from the Stanford University community participated in this study in exchange for payment. Participants were of a variety of races/ethnicities and both genders.

Materials 24 pairs of sentences were constructed which were identical for the first few words (the section underlined in 1a and 1b below) except for a critical word (italicized below). The critical words in each pair of sentences were identical except for the presence or absence of a stop at the end of a final consonant cluster:

1a. The *mast* probably lasted through the storm.

1b. The *mass* probably lasted an hour on Sunday.

These nearly identical sections (underlined above) would be ambiguous when spoken aloud if a speaker used the deleted variant of a word like *mast*. Because the deleted variant is a possibility, a listener would not be able to tell whether the word *mass* or *mast* was intended at this point in the sentence. The pairs of sentences, however, are all disambiguated by the endings of the sentences, which are more consistent with one of the interpretations of the beginning. For example, *through the storm* is more consistent with the *mast* interpretation of the beginning, and *an hour on Sunday* is more consistent with the *mass* interpretation of the beginning.

24 filler pairs were created that also contained an ambiguity that was resolved later in the sentences:

2a) While Bill hunted the deer ran into the woods.

2b) While Bill hunted the deer we made the fire.

None of these ambiguities were related to *t/d* deletion. While some of these temporary ambiguities could be disambiguated by prosody, care was taken to select recordings in which the prosody would be appropriate for both readings of the ambiguous portion. These ambiguous fillers masked the experimental sentences; when a subject encountered a temporarily ambiguous sentence, it was only an experimental sentence half the time. This prevented the experimental sentences from standing out from the fillers, as well as preventing them from exclusively constituting the most difficult sentences to understand and respond to.

In addition, 48 unambiguous filler sentences were constructed of similar length and complexity. Thus the full set of stimuli includes 24 pairs of experimental stimuli, 24 pairs of ambiguous fillers, and 48 unambiguous fillers, which were unpaired. The total number of sentences used in the experiment was 144; however, because each subject only heard one sentence from each pair, this resulted in 96 total sentences presented to each subject, half of which were ambiguous. The ratio of fillers to target stimuli was 3 to 1.

I recorded 16 Stanford graduate and undergraduate students reading all 144 of these sentences aloud. The participants were 4 African American males, 4 European American males, and 8 females of various races/ethnicities. Each sentence was preceded by a context sentence to make the participants' reading of the experimental and filler sentences more natural:

4a) I went to a new church last week that has very short services. (context)

The mass probably lasted an hour on Sunday. (target)

4b) I hope my old boat wasn't damaged by the wind last night. (context)

The mast probably lasted through the storm. (target)

Many of these speakers were recruited from the Stanford University Linguistics Department subject pool; because the subject pool did not contain many African American participants, additional African American speakers were recruited via email and paid ten dollars for their time. Recordings were digital, made in a sound-attenuated booth. Sentences were displayed to the subjects using the experimental software PsyScope (Cohen, MacWhinney, Flatt, and Provost, 1993) on a Macintosh computer.

Recordings of the male speakers were used for the target stimuli and some of the fillers. Recordings of the female speakers were used for the remaining fillers (making up half of the total clips heard by participants). Each target item was recorded once by an African American male speaker, and once by a European American male speaker. Participants were randomly assigned to hear an equal number of items recorded by

African Americans and European Americans in each face condition, so that the pairing of voice and face was equally felicitous across conditions, on average.

The speaker pictures from Experiment 1 (4 Black males and 4 White males) were used for the critical trials and one third of the fillers (24 targets and 24 fillers), while 8 females of various races/ethnicities were displayed with the other two thirds of the fillers (48). This resulted in each subject seeing a female face in half the trials and a male face in the other half of the trials. The faces and voices were paired so that each face appeared with only one voice within each subject, to maximize the plausibility of the premise that the pictures represented the speakers. Between subjects, each face appeared with one African American voice and one European American voice, so the influence of the faces can be evaluated independently of the influence of voices. Male voices always appeared with male pictures, and female voices always appeared with female pictures.

Because all subjects saw a mix of genders and races/ethnicities in the experiment, they were unlikely to be able to deduce that the experiment concerned only African American and European American males. Thus, unlike the case in Experiment 1, neither *t/d* deletion nor race was salient in this experiment.

Procedure Participants were instructed to listen to a short sound clip while looking at a picture of a face, which they were told represented the speaker of the clip. They heard the ambiguous portion of one of the sentence pairs, which contained no final stop at the end of the cluster in the target word. Although both sentences in each pair were recorded by the readers, the sound files used in this experiment were only excerpted from recordings of the sentence in each pair that never contained a final stop (e.g. the *mass* version). Participants never heard any version of the experimental sentences in which the target word was intended to contain an underlying *t/d* by the speaker, so that there are no cues in the speech stream indicating the presence of a deleted stop.

Participants then saw one of the sentence endings appear below the picture of the speaker (see Figure 5.1). For example, in one trial, participants heard:

> The [mas] probably lasted

After this clip, one of the following endings appeared on the screen:

> …through the storm.
>
> …an hour on Sunday

In half the cases, participants saw a continuation that was more likely if the ambiguous word had no final stop (e.g. *an hour on Sunday*, which was more likely if the word was *mass*), and in the other half of cases they saw the other continuation, which was more likely if the ambiguous word did have a final stop that had been deleted (e.g. *through the storm*, which was more likely if the word was *mast*). Participants' job was to assess whether the ending created a 'sensible' sentence in combination with the beginning they had heard, and response times were measured from the time the continuation appeared on the screen.

Because the ambiguous words could be interpreted as either *mass* or *mast* and thus both continuations could be interpreted as sensible, very few *no* responses were expected. Since a *yes/no* response measure was unlikely to be sensitive enough to detect an effect of race, response time was the dependent measure in this design[5]. If listeners use the relationship between race and *t/d* deletion in resolving ambiguity, then they should respond faster to the continuation that is consistent with the *t*-word (*mast*) interpretation when the purported speaker is Black than when he is White. Conversely, listeners should respond faster to the continuation that is consistent with the non-*t*-word (*mass*) interpretation when the purported speaker is White than when he is Black.

---

[5] Participants were instructed to respond 'yes' if the sentence ending was 'sensible,' which could potentially apply even to sentences like *The mast probably lasted an hour on Sunday*, which are not likely but are interpretable, and thus might fit some participants' definitions of 'sensible.' If participants came to such interpretations, this should result in slower responses, because these interpretations are difficult to reach, which will be indistinguishable from slowdowns produced by changing interpretations of the ambiguous word. Thus, either a strict or a loose interpretation of 'sensible' on the part of participants will result in the same behavior (that is, slow responses when the sentence ending does not make as much sense with their original interpretation).

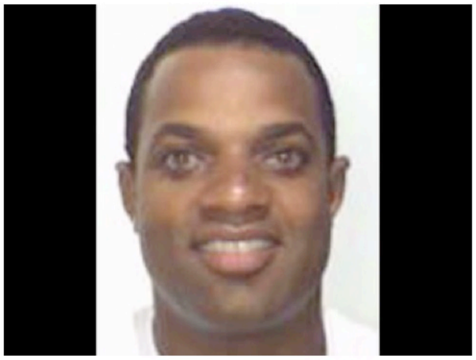

Figure 5.1. Participants in Experiment 2 saw a picture of the speaker while they listened to the ambiguous portion of each sentence. Then the sentence continuation appeared on the screen while the picture remained, and participants responded to say whether they thought the continuation on the screen was an appropriate ending for the sentence fragment they had heard through the headphones.

5.1.2 Results and Discussion

As predicted, participants responded marginally faster to the continuation that was compatible with the word whose underlying phonemic form has a *t* (the *mast* interpretation) when they saw a Black face (t1(1,39)=1.21, p=.11, t2(1,23)=1.8, p=.04). Also as predicted, they responded faster to the continuation that was compatible with the word whose underlying phonemic form does not have a *t* (the *mass* interpretation) when they saw a White face (t1(1,39)=1.75, p=.04, t2(1,23)=1.81, p=.04). This difference of differences yields the predicted significant race by *t/d* deletion interaction (F1(1,39)=5.64, p=.02, F2(1,23)=9.23, p=.006) (Figure 5.2, below).

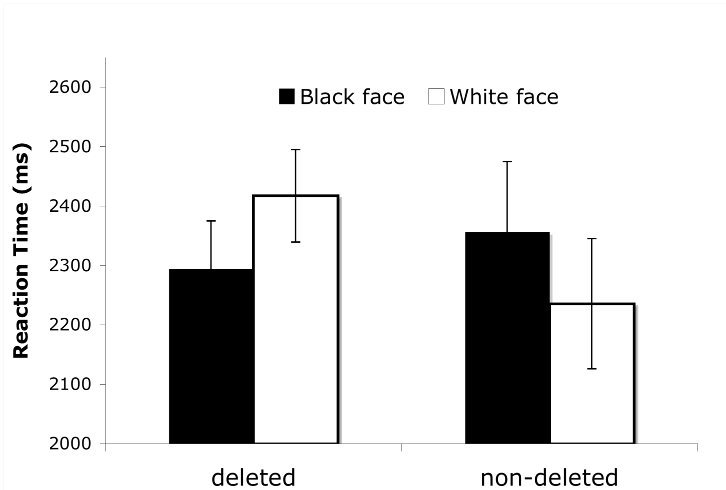

Figure 5.2. Participants in Experiment 2 responded faster to the continuation that was consistent with the deleted interpretation (e.g. mast) when they saw a Black face representing the speaker, but they responded faster to the continuation that was consistent with the non-deleted interpretation (e.g. mass) when they saw a White face representing the speaker. Error bars indicate standard errors of the means.

In addition to getting race information from the faces they saw, participants could also have made inferences about the race of the speaker based on features of the voice they heard other than *t/d* deletion. This study did not involve any assessment of how much race information was carried in each voice; however, the clips were all spoken by either a European American or an African American speaker. Thus, although it is not possible to analyze response times as a function of race information in the voices, it is possible to analyze response times as a function of the actual race of the speaker of each clip.

The actual race of the speaker whose voice listeners heard in each clip did not influence their response times, controlling for the race of the faces they saw ($F1(1,38)=1.88$, $p>.1$; $F2(1,22)=2.38$, $p>.1$). Thus, it appears that whatever race information was available in the voices did not have an influence on listeners' assessments of speaker race, perhaps because it was overwhelmed by the clear information about speaker race available in the photographs of faces.

5.2 Experiment 2b

The results of Experiment 2 are consistent with the hypothesis that listeners make use of information about sociolinguistic variation in language processing. However, interpreting this pattern of results as supporting the experimental hypothesis requires that the presence or absence of a final stop be the only important difference between the *t*-word sentence endings and the non-*t*-word sentence endings. The two sentence endings in each pair differ in several ways beyond the fact that some of them implied an instance of *t/d* deletion in the beginning of the sentence and some did not – the content of the sentence endings is different, and the register of the sentences could be different, among other things. Thus, there are many possible differences between the sentence endings that could result in one continuation being responded to more quickly than the other in each pair, beyond participants' linguistic stereotypes about *t/d* deletion.

These differences in the continuations that are unrelated to *t/d* deletion could have caused the pattern of results observed in Experiment 2a if they make the *t*-word endings more associated with African Americans, on average, than the non-*t*-word endings are. This association could be based on meaning differences in the sentences; for example, in the *mass*/*mast* pair of sentences, one sentence is about church-going, and the other is about ships. If participants associated church-going more with Black speakers and ships more with White speakers, then the association between these activities and race would predict a difference in reaction time to the two sentence endings (although in the case of these particular associations it would be the opposite of the difference found in Experiment 2a). Because results in Experiment 2a were aggregated across items, they could only be explained by this kind of association if the *t*-word sentence endings were on average more associated with African Americans than the non-*t*-word sentence endings were. However, if this were the case, then the results of Experiment 2a could be due to the content of the sentences, rather than the need to restore a deleted *t* in the *t*-word cases.

The *t*-word sentences could also be associated with African Americans due to register or dialect differences from the non-*t*-word sentences (although any differences must reside in the written sentence endings, so phonological factors cannot be involved). For example, if the *t*-word sentence endings contained lexical items or syntactic constructions that are characteristic of some variety African American English, this could cause them to be associated with speakers of that variety, who are overwhelmingly African American.

For the results of Experiment 2a to support the experimental hypothesis that race information is used to make inferences about *t/d* deletion, it is necessary to rule out alternative explanations like those discussed above for the pattern observed. To rule out these alternative interpretations, I conducted a norming experiment on written versions of the sentence pairs used in Experiment 2a. These written versions have all the same content, register and dialect differences that the sentences in Experiment 2a have, but because they are unambiguous, they do not allow for any inferences about *t/d* deletion.

5.2.1 Methods

Participants 58 native English-speaking Stanford University undergraduates received course credit for their participation in this study. These participants were a subset of the participants in Experiment 1 (Chapter 3).

Materials Materials for this experiment were identical to those used in Experiment 1, Questionnaire A. This questionnaire presented the sentences from Experiment 2a in written form in their entirety (rather than hearing the beginning but seeing only the continuation in writing, as in Experiment 2a), so that participants were exposed to the same content as in Experiment 2a, but all in the written modality. Sentences were presented in standard orthography, underneath pairs of Black and White faces (the same faces that were used in Experiments 1a, 1b, and 2a). See Figure 3.1 for an example of what participants saw.

Procedure Participants were asked to select which of the two pictures above each sentence was more likely to represent the speaker of the sentence. Each participant saw only one sentence from each pair, so that no two sentences started with the same words for each participant. Thus, the comparison between the two sentences in each pair is a between-subjects comparison.

5.2.2 Results and Discussion

There was no significant difference between the *t*-word and non-*t*-word sentences ($t(1,57)=1.05$, $p=.29$) (Figure 5.3, below). In addition, the slight numerical difference between the two lists suggests that if anything, the non-*t*-word sentences were more associated with the Black potential speakers. If this association were statistically significant, it would be in the opposite direction from the difference that would be needed to account for the results of Experiment 2a. Thus, general differences in content, register, or dialect between the two lists of sentences cannot account for the results of Experiment 2a.

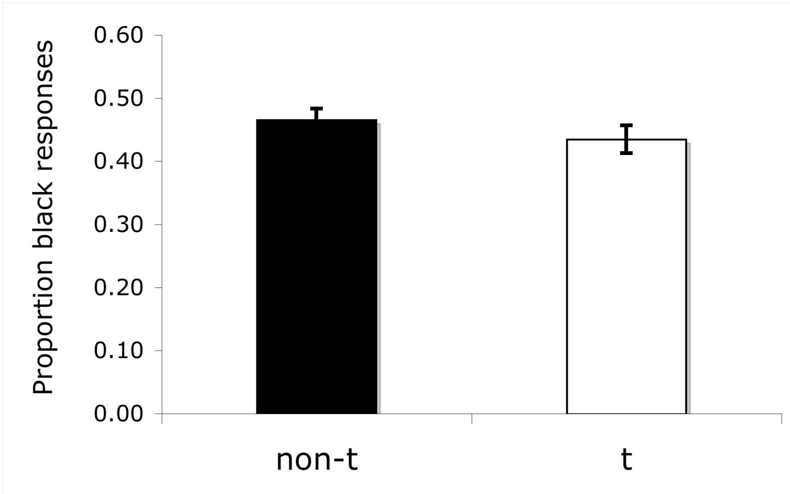

Figure 5.3. Participants in Experiment 2b (a written questionnaire) were not biased toward the Black potential speaker any differently for t words (e.g. mast) than for non-t-words (e.g. mass). Error bars indicate standard errors of the means.

5.3 Summary

Results of Experiment 2a were consistent with the experimental hypothesis that listeners use social information to resolve ambiguities, but this experiment left open the question of whether other differences in the stimuli could have caused the observed difference. The results of Experiment 2b ruled out content-, register- or dialect-based explanations for the results of Experiment 2a, confirming that expectations about *t/d* deletion are the likely explanation for the observed differences in behavior. Together, Experiments 2a and 2b provide evidence that listeners combine their knowledge of sociolinguistic variation, social information from the scene, and auditory information from the speech stream to construct an interpretation of the speech they have heard. This is consistent with an optimized view of language comprehension, in which listeners make use of all sources of predictability to which they have access when trying to solve the difficult problem of speech perception.

Chapter 6

Representing Sociolinguistic Knowledge

Experiments 1a and 1b demonstrated that people associate deleted *t/d* more strongly with Black speakers than with White speakers, and that they develop these associations through experience with *t/d* deletion. Experiments 2a and 2b demonstrated that listeners make use of information about the speaker's race from the context to resolve ambiguities based on *t/d* deletion. The results of these experiments, taken together, indicate that information about speakers is included in listeners' mental representations of linguistic forms. But what is the nature of these mental representations? How could this influence of social information be incorporated into a model of speech perception? This chapter discusses some current models of speech perception and how they could accommodate the influence of social information, and then presents two experiments constraining the way this influence should be incorporated into such a model. The first experiment investigates whether a categorical phonetic level of representation is necessary for listeners to make inferences about *t/d* deletion based on characteristics of the speaker. The second experiment investigates whether correspondences on the level of sub-lexical chunks of phonological material between incoming tokens and previously experienced tokens are sufficient to allow listeners to make inferences based on these previous experiences.

6.1 Effects of context in models of speech perception

Some existing models of speech perception have already been adapted to account for the influence of context on various tasks such as word recognition. For example, speaker-specific effects have been found in several speech perception tasks (e.g., Goldinger, 1998; Luce and Lyons, 1998;). Speaker specific effects are ones in which a word spoken by a speaker the listener has heard before is recognized more quickly than one spoken by an unfamiliar speaker. Such effects suggest that acquiring sufficient experience (which may not be much) with an individual speaker allows listeners to tune their perception to qualities of that speaker. Any model in which the categories used in speech perception are sensitive to prior experience can predict these kinds of effects.

One class of models that has been used to account for speaker-specific effects on speech perception is exemplar models, in which listeners have detailed episodic memory traces of linguistic experiences that include details not only of the acoustic signal they perceived, but of many aspects of the context in which the signal was perceived (Goldinger, 1996; Johnson, 1997; Pierrehumbert, 2000). The motivation for importing such models from perception and categorization in general to speech perception was to account for detailed phonetic knowledge that speakers and listeners have about specific words in their lexicons; exemplar models can account for the effects of lexical frequency on phonetic reduction, for example (Bybee, 2000). The model that has been invoked to account for many speaker-specific effects can also account for effects on the level of groups of speakers (and has been invoked to do this as well [e.g., Hay, Warren, and Drager, 2006]). In exemplar models of phonological knowledge, social information about speakers can be available to the listener by virtue of indexing of the tokens of past experiences that are stored by the listener.

An exemplar model in which the exemplars that are stored are indexed according to characteristics of the speaker can easily account for the results of Experiment 2. If listeners have socially indexed tokens, then when they hear someone utter [mæs], it will correspond to some of their stored exemplars of *mass* and some of their stored exemplars of *mast*. If they believe that the speaker is Black, then they will preferentially activate tokens of these two words that are indexed to Black speakers, because these tokens have more in common with their current input. In this case, a greater proportion of the tokens of [mæs] will correspond to *mast* than if the speaker were White, since [mæs] is a more frequent pronunciation of that word for Black speakers than White speakers. If, by contrast, the listener believes that the speaker is White and preferentially activates tokens that are indexed to White speakers, then more of the tokens of [mæs] will correspond to *mass*, because deletion is less common among White speakers. The more stored tokens that correspond to the representations cued by the input, the faster the input will be recognized. Thus, the social indexing of the exemplars clearly predicts the effect found in Experiment 2.

Including social indexing in an exemplar model requires that the detailed episodic traces of linguistic experiences include details about the speaker. The stored details of these experiences allow listeners to associate aspects of linguistic form with characteristics of speakers. Relationships between social characteristics and sociolinguistic variables are thus generalizations across stored tokens. However, the architecture of an exemplar model provides potential limits to the types of inferences that listeners might be able to make about future speaker behavior. Specifically, an incoming token must correspond to some previous token in some way in order to activate details of the previous token as it was experienced by the listener. This aspect of the model makes a testable prediction: social information should only influence the perception of tokens that correspond to previously experienced types. But what constitutes a type in exemplar theory?

In a strict version of exemplar theory, where tokens are episodic traces of previously experienced exemplars, the basic unit of exemplar storage and the abstractions that can be made over these basic units are of crucial importance in making behavioral predictions based on the model. In fact, the potential flexibility of the process of abstraction across the space of stored tokens in an exemplar model can make it difficult to determine at what level the exemplars are actually stored. While there is no universal consensus on this matter, Johnson (2005) has suggested that the word may be the unit at which exemplars are stored, because words are more accessible to speaker/listeners than sounds. The unit of exemplar storage is important because, as Johnson points out, in exemplar-based models instances of language are stored as they occur, without any abstraction in the storage process. Thus, if words are the basic unit of exemplar storage, any categories beyond the word are not stored in the system. However, Johnson is quick to point out that generalization and abstraction are still possible, as a response of the aggregated exemplars. This leaves the range of behaviors that could be accounted for by an exemplar-based model fairly unconstrained, depending in great part upon what types of generalization are proposed to emerge from the stored word-level exemplars.

Experimental results have provided some information about what levels of abstraction should be a part of a successful model of speech perception. For example, while it is theoretically possible for lexemes to map directly onto acoustic or articulatory properties, there are empirical results suggesting the need to posit an intervening phonological level of representation. Pierrehumbert (2006), in motivating what she calls a Hybrid Model incorporating aspects of generative phonology with aspects of exemplar theory, suggests that even an exemplar approach to phonological categories requires a phonological level of representation to account for the relationships between phonotactics and lexical neighborhood densities demonstrated by Vitevich and Luce (1998). While these two characteristics of words are highly correlated, with words that have many lexical neighbors also having high-probability phonotactics, it is possible to vary them independently in an experiment. Vitevich and Luce found that words with high probability phonotactics were recognized faster than words with low probability phonotactics, when they were equated for the density of their lexical neighborhoods. However, words with many lexical neighbors were recognized more slowly than words with few lexical neighbors, once these words were equated for the probability of their phonotactics. As Pierrehumbert (2006) points out, for these two factors, which are highly correlated, to influence behavior in opposite ways, they must operate on different levels of representation. Indeed, like most theories of phonology, exemplar theories now tend to include the possibility of such a categorical phonological representation, in their case derived through a process of abstraction and generalization across the space of exemplars encountered by a listener.

Given that a nearly infinite set of abstractions is theoretically possible in an exemplar model of speech perception, specifying the kinds of abstraction that actually occur depends on empirical results that support a particular type of abstraction, such as the Vitevich and Luce (1998) study. Their results depend on the existence of word-level units and phoneme-level units in speech perception; what other levels of abstract representation does a model of speech perception require?

In Experiment 2, listeners' representations of speech were probed by asking them to judge the felicity of sentence endings that were consistent with only one meaning of the ambiguous string they heard. Listeners were never asked what word they thought they had heard, never heard a target word pronounced with a final *t* or *d*, and never encountered an orthographic representation of the word or any other signal that any of the target words had a phonetic or phonological final stop. The results of this experiment showed that listeners came to different interpretations of the same acoustic signal based on the perceived race of the speaker. However, because these were responses to sentence endings that were consistent in meaning with one interpretation or the other, these responses only give information about the representation they formed on the level of semantic meaning – the phonetic and phonological representations listeners formed during this task are unknown.

If one accepts the need for at least one categorical phonological level of representation, there are still many possible systems of levels of representation. The relationship between phonemic representations and acoustics or articulation is far from direct; as discussed in Chapter 1, there is a double-dissociation between categories of sounds and acoustic cues, such that the same cues can indicate different sounds in different contexts, and the same sounds can be manifested by different cues in different contexts. Several solutions to this problem have been posited, with varying representational components.

Sapir produced one of the first models of how phonemic structures of words could correspond to the sounds speakers actually produce. Sapir's *Item and Process* model contains one level of representation mediating between phonemic representations and acoustics/articulation. This system, which employs a categorical phonetic level of representation in addition to the phonological representation, was the dominant one in the early days of generative phonology. The phonetic representation was derived from the phonemic representation by a set of ordered phonological rules. Many phenomena from a variety of languages have been modeled according to these general principles very successfully, perhaps inspired by Chomsky and Halle’s seminal work in the area, *The Sound Pattern of English* (1968). Despite its long list of successes, this model has since come under attack for positing unjustified and unnecessary theoretical machinery: “…when one probes with questions of what exactly the phonological and phonetic representations are, and why language should involve the transformation of one into the other, no very convincing answers are forthcoming” (Fraser H., 1997, p. 110).

Aside from objections simply based on lack of justification, some theorists have pointed out that a workable alternative involving only one level of representation would be a priori preferable, due to its simplicity; Harris (2003) points out that the null hypothesis should be that phonological categories map directly to articulation and auditory perception. Motivated primarily by parsimony, theories involving direct relationships between phonemic representations and acoustics/articulation have been developed. Monostratal versions of Optimality Theory (Kirchner, 1997), for example, may locate the phonetic details of speech in the phonological representations and eliminate the phonetic level altogether. With constraints doing much of the work in Optimality Theory that was done by representations and derivations in previous models, the usefulness of a categorical phonetic level of description was seriously called into question. However, despite the appeal of theoretical parsimony, distinctions between phonetic and phonemic levels of representation continue to be implicated by experimental results in speech perception.

McClennan, Luce, and Charles-Luce (2003) discovered processing differences between words that are phonetically ambiguous but phonologically distinct and those that are both phonetically and phonologically distinct. They found that carefully and casually produced tokens of words with alveolar consonants that undergo flapping in casual speech prime each other in a long-term priming paradigm, but carefully and casually produced tokens of words with non-alveolar consonants do not. They suggest that the words with alveolar medial consonants, which can become ambiguous due to flapping (e.g., *Adam*, *atom*), are restored during perception to an inferred phonetic representation (which they call a *surface representation*) based on the phonology of the lexical item, such that the casually produced words match the carefully produced words on this phonetic level of representation. By contrast, the words with non-alveolar medial consonants (e.g., *bacon*) are never ambiguous and do not require restoration, so the differences in production of the casually and carefully produced words are retained in this phonetic level of representation. The *surface* level of representation they posit cannot be phonemic, because the words with non-alveolar consonants are phonemically identical, and yet they do not produce any long-term priming when their pronunciations do not match.

Despite the theoretical motivation for doing away with categorical phonetic representations, and the existence of the theoretical machinery to account for much of speech perception without them, the number of empirical results that seem to depend on such a level of representation is large enough that the possibility of its existence must be taken seriously. Thus, the effects discovered in Experiment 2 could arise in at least two different ways. First, listeners might form the same phonemic representations of words for all speakers, but map them onto different phonetic representations, which would require a separate phonetic level of representation. Alternatively, listeners might form the same phonemic representations of words for all speakers, but map them directly onto acoustic information they get from the speech signal without any categorical phonetic level of representation. This would allow for the results of Experiment 2 without differences on any categorical level of representation. Is there a categorical level of representation on which listeners form different representations for black and white speakers? Experiment 3 takes advantage of the properties of visual word recognition to address this question.

6.2 Experiment 3a

There is evidence from a variety of tasks suggesting a dual-route model of visual word recognition, in which orthographic representations are also cues to phonology (Luo, 1996; Van Orden, 1987; 1991). Lesch and Pollatsek (1998) showed that phonological codes cued by orthography are not directly mapped onto lexical items but are assembled from phonemic components cued by the orthography. Participants in this study had more trouble judging pairs like PILLOW-BEAD, which contained false homophones to related words (like BED), to be semantically unrelated than orthographic control pairs like PILLOW-BEND. BEAD is considered a false homophone of BED because the string *–ead* in English is sometimes pronounced the same way as the *–ed* in BED. Lesch and Pollatsek interpret these results as indicating that participants were constructing phonological codes out of phonemic components, because the orthography of BEAD gives cues both to a code that sounds like BED and one that does not.

In fact, orthographic cues can even activate representations that conflict with a speaker's normal pronunciation of a word (Taft, 2006). Speakers of a non-rhotic dialect (Australian English) had trouble judging that the pseudohomophone CAWN is homophonic with CORN, even though in their dialect these words would be pronounced the same way. This suggests that their representation of CORN contains a phonemic /r/, despite the fact that their pronunciation of this word does not contain any rhoticity.

While some of these results suggest that phonological representations are activated during visual word recognition, there is evidence that phonetic representations are also activated during this process. Birch, Pollatsek, and Kingston (1998) investigated the role of phonemic and phonetic codes in visual word recognition, and found that both phonemic and phonetic representations of visually presented pseudowords were involved in homophone judgment tasks. However, they only found evidence for the use of phonetic representations, and not phonemic representations, in a lexical decision task on pseudohomophones, suggesting that the codes involved in visual word recognition are dependent on the task. For tasks that do not explicitly involve comparing sounds of words, a phonemic level is not implicated by their findings, while a phonetic level of representation seems to be involved in a variety of tasks. These findings do not indicate that a phonemic level of representation does not exist or is not involved in visual word recognition; rather, they suggest that a separate, categorical phonetic level of representation is also involved in this process.

If reading a word activates phonetic and sometimes phonological representations of the component sounds of the word, then showing an orthographic representation of a word which includes a *t* (e.g. *mast*) should also activate a representation of a [t] in the listener's mind, even if listeners did not have a [t] in their representation of the word prior to seeing it written. By contrast, showing an orthographic representation of a word which does not include a *t* should not activate a representation of a [t]. Experiment 5 takes advantage of these properties of visual word recognition to investigate whether listeners form different phonetic representations of words that are ambiguous due to potential *t/d* deletion based on the race of the speaker.

6.2.1 Methods

Participants Thirty-nine native English speakers from the Stanford University community participated in this study in exchange for payment. All participants had lived in the United States for at least 18 years. Participants were of a range of races/ethnicities and both genders, and most were between 18 and 22 years old.

Materials Some materials for this experiment were adapted from Experiment 2. Target items were 24 sentence beginnings each containing a word that is ambiguous between a word ending in a consonant cluster with a *t* or *d* in the final position (such as *mast*) and a word that is identical save for the absence of the *t* or *d* (such as *mass*). The sound files were excerpted from recordings of entire sentences read by Stanford graduate and undergraduate students, who were paid for their time. Participants heard excerpts from sentences that never contained an underlying final stop (i.e., they heard sentences in which speakers intended to say *mass* but never sentences in which speakers intended to say *mast*). Thus, participants never heard any version of the experimental sentences that contained an underlying *t/d*, so that there were no cues in the speech stream indicating the presence of a deleted stop.

Each target item was heard spoken by an African American speaker by half the subjects and spoken by a European American speaker by the other half of the subjects, and the race of the actual speaker was crossed with the race of the pictured speaker.

Having both types of speakers in the Black and the White face conditions prevented one race condition from being generally more felicitous with the voices heard than the other. However, the acoustic cues to race/ethnicity (other than *t/d* deletion) available in each clip varied naturally, and were not controlled. Listeners could potentially have been influenced by cues to the race of the speaker that were present in the audio clip; in analysis, the actual race of the speakers was used as a proxy for cues to race in the speech stream. If cues from the speech stream influence listeners' reactions, they should do so in the same way that cues from the pictures are predicted to do.

Forty-eight similar fillers were constructed that also consisted of only the beginning portion of a sentence. One word was selected from each sentence beginning to serve as the false target. In addition, sixteen similarly structured fillers were created that contained words that could be subject to *t/d* deletion without creating ambiguity. For example, the word *fast*, when subject to *t/d* deletion, becomes [fæs], which is not a word in English. These sentences were recorded by a non-naïve speaker[6], who was instructed to produce the words without a final stop. As with the first 48 fillers, the beginning portions of these sentences were used, with the words with deleted final stops serving as false targets. The purpose of these fillers was to make the overall tone of the experiment more casual, and to encourage participants to believe that the speech they were hearing might contain informal variants like deleted *t/d*. However, these

[6] This speaker had to be aware of the focus of the experiment because he needed to specifically avoid producing audible final consonants in the ending clusters of the crucial words. These productions may have contained cues to an underlying /t/, but this would not interfere with their function of giving participants reason to believe that *t/d* deletion was compatible with the speech situation of the speakers in the experiment.

sentences were produced by a different speaker so that participants did not have any a priori reason to believe that a particular speaker of the target words would or would not engage in *t/d* deletion.

In addition to the 64 real word fillers, this experiment included 24 ambiguous nonce word fillers (see Appendix for a list of all stimuli used in the experiments). Like the ambiguous real words, these nonce words were ambiguous between a word with a deleted *t* or *d* at the end of a final consonant cluster, and a word that is identical except for the final stop in the consonant cluster. These words were all recorded by naïve speakers who were reading the non-*t* version of the word (e.g., *frass*), so there were no cues to a stop in the acoustic signal.

Each nonce word was paired with one of the ambiguous sentence beginnings, creating a phrase identical to one of the real phrases except for the nonce word replacing the target word:

The *frass* probably lasted…

The *frast* probably lasted…

Each participant heard each carrier phrase twice – once with a real word, and once with a nonce word. These two instances of the carrier phrase were always in different halves of the experiment. In total, there were 24 target items and 88 fillers; 40 of these could have been interpreted as containing a t/d deletion.

Each sound clip was presented with a photo of a purported speaker; photos were identical to those used in Experiment 2, with the addition of one male who was of East

Asian descent (See Appendix for photo), who was matched with the voice that produced the 16 fillers containing unambiguous *t/d* deletions. In all, there were 9 photos of males and 8 photos of females used in this experiment.

Procedure Participants were instructed to listen to a short sound clip while looking at a picture of a face, which they were told represented the speaker of the clip. They heard the ambiguous portion of one of the sentences, which contained no final stops, e.g., *The [mæs] probably lasted*. While they were listening to the clip, participants saw the words in the phrase they were hearing below the picture of the speaker, with one of the words replaced by an underlined space (See Figure 6.1 below). This phrase appeared at the beginning of the trial, at the same time that the clip began. Participants then saw either the *t* version or the non-*t* version of each word appear below the picture of the speaker; the word appeared after the clip was finished playing, so that participants had already finished processing the auditory stimuli by the time the written word appeared.

Participants pressed Y to indicate that they believed the word on the screen was the word they had heard (that went in the blank), and N to indicate that they believed the word on the screen was not the word that they had heard. Response times were measured from the time the target word appeared on the screen. In approximately half of the trials participants were presented with plausible transcriptions of the word in the audio clip, and in the other half they were presented with implausible transcripts of this word, although the target items were all presented with a plausible transcription of the target words.

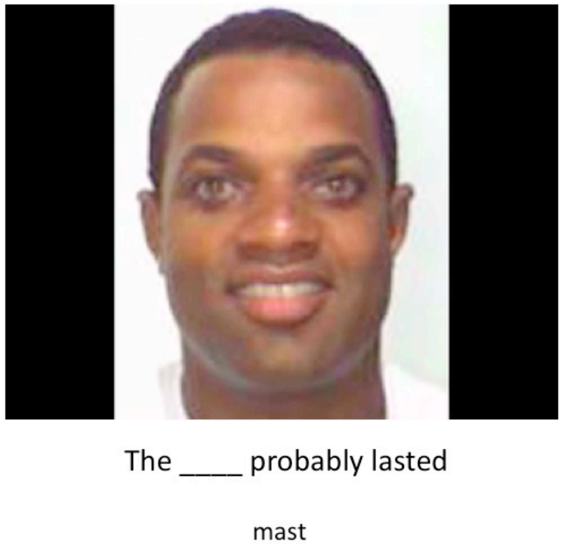

Figure 6.1. Participants in Experiment 3a saw a picture of the speaker while they listened to the ambiguous portion of each sentence. The words of the ambiguous portion of the sentence appeared on the screen while the clip was played, with an underlined section replacing the critical word. Then either the t or non-t word appeared on the screen, and participants responded by indicating whether they thought this word was the missing word in the phrase on the screen.

Each participant either saw Black faces matched with non-*t* words and White faces matched with *t*-words, or Black faces matched with *t*-words and White faces matched with non-*t* words, creating a between-subjects design. As in Experiment 2, each voice was presented in half the trials paired with one Black face and in the other half of the trials paired with one White face (between subjects), so that the race of the speaker and of the person pictured were crossed. Each subject heard each voice paired with only one picture, to increase the likelihood that the participants interpreted the people pictured as the speakers of the clips.

6.2.2 Predictions

The results of Experiment 2 indicate that listeners formed different lexical representations for different speakers – that is, their judgment of the word they think they have heard is influenced by the speaker's perceived race. Because these words contain different phonemes, listeners should also have formed different phonological representations for different speakers. If these phonological representations map directly onto auditory perception, then the results of Experiment 3a should be parallel to those of Experiment 2: listeners should respond faster to the *t*-word (*mast*) when they have seen a Black speaker, and they should respond faster to the non-*t*-word (*mass*) when they have seen a White speaker.

Alternatively, these phonological representations may map onto categorical phonetic representations. In this case, results may diverge from those of Experiment 2. If participants are presented with a written token of *mast*, but they have interpreted the *t*-less token they heard as a token of *mass* (which should happen more often if the speaker is White), then the phonetic representation cued by the orthography will conflict with participants' already formed representations, regardless of the race of the speaker (since no speaker of any race pronounces *mass* with a *t*). However, if participants have interpreted the *t*-less token they heard as a token of *mast*, (which should happen more often when the speaker is Black), the phonetic representations they form upon hearing the word may be different.

Although the speech stream does not contain a *t*, it may sometimes be included in listeners' phonetic representations of the word they have heard because these representations do not correspond perfectly to what is in the speech stream. Thus, the race of the pictured speaker could influence the phonetic representation listeners form based on constant phonetic information, because listeners take into account the likelihood of hearing a *t* when they determine whether or not they have heard one, as happens in phoneme restoration (Warren, 1970). In this case, participants should form representations without a [t] more often when the speaker pictured is Black, and they should form representations with a [t] more often when the speaker pictured is White. These likelihoods are based on participants' beliefs about how often different types of speakers will produce deleted and non-deleted tokens. If listeners form different phonetic representations for speakers of different races, then participants should respond more slowly to the word *mast* if they have seen a Black speaker than if they have seen a White speaker. This prediction may seem surprising at first, because it is exactly the opposite of the prediction based on the results of Experiment 2.

Predictions for listeners' reactions to the orthographic representation of the non-*t* word are also different from the findings from Experiment 2. If participants see an orthographic representation of the non-*t* word, such as *mass*, this will be consistent with their phonetic representation if they have interpreted the word as *mass*, no matter what type of speaker they have seen pictured. However, if they have interpreted the word as *mast*, then the phonetic representation cued by the orthography will be consistent with the representation they have already formed more often when the speaker is Black. This suggests that participants should respond faster to words like *mass* if the speaker is Black than if the speaker is White.

To summarize, the two possibilities make opposite predictions for this task: if participants map phonology directly onto auditory input, seeing the word *mast* should be more consistent with seeing a Black speaker and seeing the word *mass* should be more consistent with seeing a White speaker (as in Experiment 3), but if participants form different categorical phonetic representations for different speakers, seeing the word *mast* should be more consistent with seeing a White speaker and seeing the word *mass* should be more consistent with seeing a Black speaker.

6.2.3 Results and Discussion

Overall, participants responded "yes" 78% of the time to target words. Both "yes" and "no" trials were included in the analysis, to avoid uneven numbers of items in each condition. One item was discarded due to improper balancing[7], leaving 23 items present in the analysis.

Listeners responded faster to the non-*t* word (e.g. *mass*) when the speaker pictured was Black than when he was White, but they responded faster to the *t*-word (e.g. *mast*)

[7] The item that included the pair of words *war* and *ward* was not seen in all conditions, so responses to this item were excluded from all analyses.

when the picture showed a White speaker than when it showed a Black speaker ($F1(1,74)=2.32$, $p=.13$, $F2(1,22)=4.56$, $p=.04$) (see Figure 6.2 below). These results are consistent with the hypothesis that listeners have different categorical phonetic representations for different speakers, such that the phonetic representation listeners form of the *t*-words is different depending on the social characteristics of the speaker.

The actual race of the speakers of the clips also influenced reaction times, independent of the race information from the pictures. Listeners responded faster to the non-*t* words when the actual speaker of the clip was African American than when he was European American, and they responded faster to the *t*-word when the speaker of the clip was European American than when he was African American ($F1(1,74)=2.9$, $p=.09$, $F2(1,22)=9.38$, $p=.006$), consistent with the results based on the race of the pictured purported speakers.

Listeners formed different representations of the ambiguous *t/d* words depending on the race of the speaker. When they believed the speaker was Black, they formed more *t*-less representations than when they believed the speaker was White, reflecting the way Black and White speakers tend to produce *t*-words; this effect was produced despite the fact that listeners actually heard the same acoustic input in all cases, suggesting that the difference observed was in the way listeners categorized that acoustic input on the phonetic level. These results support the existence of a categorical phonetic level of representation in speech perception, and suggest that in a model of speech perception in which categories are emergent from the input, phonetic categories play a useful role.

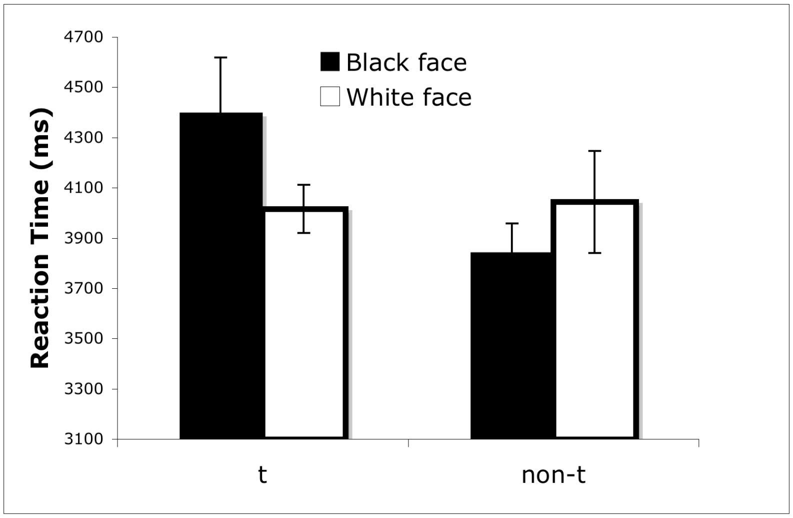

Figure 6.2. Participants in Experiment 3a responded faster to the t-word (e.g. mast) when they saw a White face representing the speaker, but they responded non-significantly faster to the non-t word (e.g. mass) when they saw a Black face representing the speaker. Error bars indicate standard errors of the means.

Providing an exemplar-based account of the results of Experiment 3a is quite straightforward. Seeing a word written with an orthographic *t* activates the phoneme /t/ and through it, the phonetic category [t]. I postulated that this resulted in faster responses to White speakers in these cases because listeners had been influenced by their expectations of White speakers into constructing a phonetic representation of what they had heard that was more likely to have a phonetic [t] than the representation they constructed in the Black speaker condition, even though the acoustic input was identical in the two conditions. This is a plausible mechanism for the effect observed. However, an exemplar model can provide another mechanism for this effect.

Seeing a White face, according to an exemplar model, may activate to some extent all utterances ever made by a White speaker in the listener's experience. Assuming that Black and White speakers produce approximately the same number of phonemic /t/s, the fact that White speakers do less *t/d* deletion suggests that they on average produce more phonetic [t]s than Black speakers. Thus, the phonetic [t] activated by the orthography of the word *mast* may be more consistent with a White speaker, in this model, simply because White speakers produce more [t]s in general.

While a phonetics-level explanation can account for participants' responses to seeing a *t*-word, this explanation does not seem to apply to the cases where no orthographic *t* was seen. When the word *mass* is seen, it activates the phonetic representation of this word, [mæs], which matches more exemplars spoken by Black speakers, because not only their productions of *mass* match this, but also more of their productions of *mast*. While it may also be the case that, parallel to the explanation of the *t*-word effect above, overall Black speakers produce more words without a [t], it is hard to see how the lack of [t] could be represented in a way that would spread activation to other words that do not have a [t]. Thus, this phonetic representation is more consistent with a Black speaker, but only because of factors related to tokens of these specific words – not because of general tendencies over exemplars of phonemes or phones. The fact that an exemplar theoretic account of this effect seems to depend on lexical representations raises another question about the abstractions that are necessary in an exemplar model of speech perception: are effects that cannot be accounted for by phonological or phonetic representations restricted to words that listeners have stored tokens of? Or can these effects extend to novel words through abstraction over an intermediate level of representation, the sub-lexical chunk?

6.3 Experiment 3b

If the unit of exemplar storage is the word, and there are no categorical levels of representation between this level and the phonological level, then listeners should not make inferences about how different speakers would pronounce words for which they do not yet have a lexical entry, and of which they have not yet stored any traces. Alternatively, however, if abstraction below the level of the word but above the level of the phoneme is possible, social information could influence the perception of never-before-heard words, as long as these words contain some previously heard items at a sublexical level.

Many potential effects of social characteristics on nonce words (especially those involving vowels) could rely on phonological representations. For example, for an American, hearing the nonce word [fras] could activate the vowel (ah), which is also heard in *frog*, if the speaker of the nonce word were American. If the speaker of the nonce word were British, however, the same acoustic information might be mapped onto the vowel (ae), which is also heard in *glass*. This would be based on the American listener's knowledge of the different correspondence between phones and phonemes in British and American English.

In the case of *t/d* deletion, however, having nonce words contain familiar English phonemes is not sufficient, because it involves the complete deletion of a phoneme. The nonce word [fræs] cannot activate the phonemic structure /fræst/ by a simple phone to phoneme mapping, because there is no phone that corresponds to the phoneme /t/. Thus it must be on the level of a sublexical chunk that this correspondence exists, if it does. That is, the chunk [æs] could correspond to the phonemic structure /æst/ by virtue of listeners' experiences with other words containing this phonemic structure (e.g., *mast*, *fast*, etc.) in which this structure is sometimes phonetically realized as the chunk [æs].

The correspondence in their experience between the phonetic chunk [æs] and the string of phonemes /æst/ could allow listeners to sometimes interpret [fræs] as /fræst/, even if they never get any acoustic cues to a /t/. Once listeners have interpreted this phonetic information as /fræst/, the same reasoning could apply that applied with the real words. Seeing a Black speaker would cause listeners to assign the phonetic representation [fræs] to this phonemic representation more often, while seeing a White speaker should cause listeners to assign the phonetic representation [fræst] to this phonemic representation more often, because of listeners' expectations regarding how different speakers will pronounce these words. Thus, the orthographic representation *frass* will be more consistent with seeing a Black speaker, and the orthographic representation *frast* will be more consistent with seeing a White speaker.

In Experiment 3b, listeners hear nonce words that might have been subject to *t/d* deletion while seeing faces of White or Black speakers, using the same paradigm that was used in Experiment 3a. Results of this task will be compared to the results of Experiment 3a to determine if never-encountered words are subject to the same effect of perceived race of the speaker that real words are. If the perception of novel words that are ambiguous due to potential *t/d* deletion is influenced by speaker race, then listeners must be relying on abstractions across sub-lexical chunks of phonetic material to make inferences about incoming tokens.

6.3.1 Methods

Participants Thirty-nine Stanford University undergraduates were paid or received course credit in exchange for their time. This experiment was conducted in the same session as Experiment 3a, so participants were shared between these two experiments.

Materials Materials for this experiment were identical to those used in Experiment 3a, with the exception that the nonce word fillers in Experiment 3a were the target items in Experiment 3b, and the real word targets from Experiment 3a were fillers in Experiment 3b. These nonce words all had a similar structure to that of the real words, with consonant clusters that could be subject to *t/d* deletion, paired with the words that would be ambiguous with them after deletion (e.g. *stip*/*stipt* or *cliss*/*clist*). The nonce words occupied the same places in the same phrases that the real words occupied in

Experiment 3a. Each carrier phrase was used twice, once with a real word and once with a nonce word, in separate blocks of the experiment. Other filler items were shared between the two experiments.

Procedure The procedure for Experiment 3b was identical to that of Experiment 3a. Both nonce words and real words were embedded in phrases and listeners were asked to identify them by saying whether or not the word they saw on the screen was the word they had heard. Reaction times were measured from the time the target word appeared on the screen. The experiment lasted between ten and twenty minutes for each participant.

6.3.2 Results and Discussion

As in Experiment 3a, "yes" and "no" responses were analyzed together to prevent uneven numbers of observations in the conditions. Three items were discarded due to improper balancing, leaving 21 items in the analysis[8]. Participants responded to nonce words with a final stop faster when the picture showed a White speaker than when the picture showed a Black speaker, and they responded to nonce words with no final stop faster when the picture showed a Black speaker than when it showed a White speaker ($F1(1,74)=3.01$, $p=.09$, $F2(1,20)=32.24$, $p=.00001$, see Figure 6.3 below). While this interaction is only marginally significant by subjects, it is highly significant by items,

[8] The nonce word pairs *stip/stipt, biss/bist*, and *kron/krond* were not seen by all participants in all conditions, so they were removed from the analysis.

and the interaction in nonce words was not significantly different from the interaction found in Experiment 3a (F1(1,74)=.06, p=.80, F2(1,42)=.68, p=.41) when participants responded to real words.

Also parallel to Experiment 3a, the actual race of the speaker whose voice listeners heard influenced reaction times. Nonce words with a final stop were responded to faster when the speaker of the clip was European American than when he was African American, and nonce words with no final stop were responded to faster when the speaker was African American than when he was European American (F1(1,74)=2.95, p=.09, F2(1,20)=23.23, p=.0001). This interaction also did not differ from the interaction found in Experiment 3a (F1(1,74)=.004, p=.95, F2(1,42)=.004, p=.95).

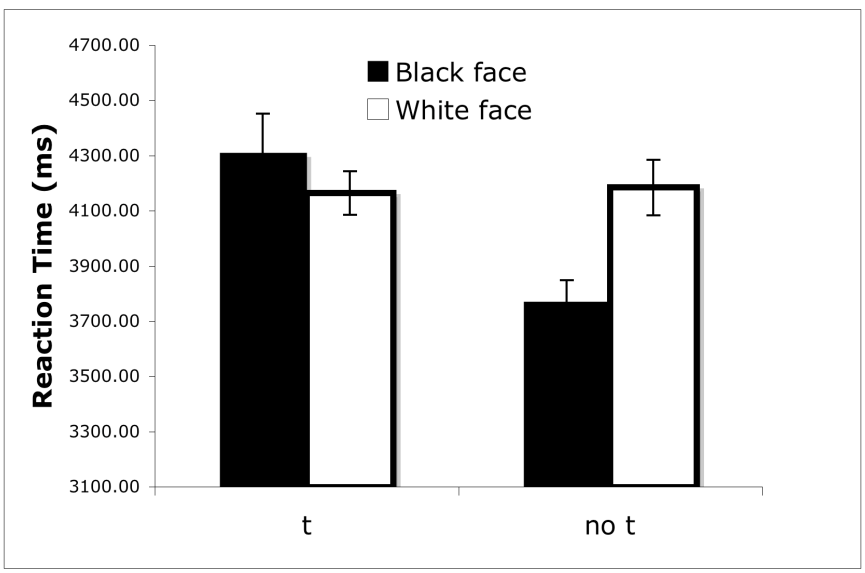

Figure 6.3. Participant responses in Experiment 3b showed the same interaction between race of the pictured speaker and word type with nonce word stimuli that they showed with real word stimuli in Experiment 3a. Error bars indicate standard errors of the means.

The apparent reliance on the exemplars of specific words to account for some of the results of Experiment 3a was consistent with Johnson's (2005) suggestion that the word is the basic unit of the exemplar model. However, this feature also makes the results of Experiment 3b very difficult to account for. In Experiment 3b, the results of Experiment 3a were replicated using nonce words. This cannot be accounted for based on matching the stimulus to previously experienced tokens of the same type, because the nonce words were specifically constructed to be word-like strings participants had likely never heard before, and thus had no exemplars of, at least on the level of the word. If a version of exemplar theory can account for this data, then it will need to include abstraction on a sublexical level.

In Experiment 3b (as in previous experiments), there were no cues to an underlying /t/ in the speech stream. In half the cases, participants saw an orthographic *t* (e.g., in the word *frast*). In these cases, as in Experiment 3a above, exemplars of real words spelled with a *t* become active, and most of these words have exemplars in which the /t/ was pronounced. Thus seeing the word *frast* activates a phonetic [t]. Although the listener has no exemplars of either *frass* or *frast* to rely on, *frast* could still be more consistent with a White speaker, because White speakers are more likely to produce a phonetic [t] in general. However, in the other half of the cases, participants did not see an orthographic *t*. The interaction in nonce words was largely carried by these cases (see Figure 6.3, above), so the difference by race in this half of the data must be accounted for. In Experiment 3a, this effect was accounted for because for Black speakers, the [t]-less phonetic string (i.e., [mæs]) was consistent with both the /t/-less word (i.e., *mass*) and many of the exemplars of the word that did contain an underlying /t/ (i.e., *mast*). For White speakers, fewer of the exemplars of these words containing an underlying /t/ matched this string phonetically. However, the “words” participants encountered in Experiment 3b were words they did not have any exemplars of at all. Thus, this mechanism is not available to account for participants’ behavior.

While it is true that listeners do not have any stored exemplars of the word *frass*, it is possible to imagine that they might have stored exemplars of chunks in between the level of the word and the level of the phoneme. In fact, one of the benefits of exemplar theory is that it is in principle compatible with any number of levels categorization, because all categories emerge through a process of abstraction over the exemplars. If exemplars of the chunk [æs] can be abstracted from the input, then *frass* will activate this chunk, which through spreading activation will activate all sequences of phonemes that were realized with this phonetic chunk, including exemplars of both /æs/ and /æst/. Because all speakers produce all of the /æs/ tokens as [æs], but Black speakers produce more of the /æst/ tokens as [æs], a greater number of consistent exemplars will be activated when the speaker is Black than when the speaker is White. This would produce an RT speedup for the same reason that reading a more frequent word is faster than reading a less frequent one.

6.4 Summary

Experiment 3a demonstrated that listeners form different phonetic representations of words that could have been subject to *t/d* deletion for speakers of different races. Experiment 3b demonstrated that the influence of social information on speech perception extends to novel words, even when no correspondence on the phonological or phonetic level is possible, indicating that listeners must be able to generalize across instances of sub-lexical chunks of phonetic material. If an exemplar model (or any kind of model) of speech perception is to account for the results of Experiments 3a and 3b, it will need to include abstractions not only at the levels of the word and phoneme, but also at the levels of the phone and the sub-lexical chunk. In the case of an exemplar model, these abstractions can emerge from patterns in listeners' detailed episodic traces of speech events.

The fact that phonetic categories and sub-lexical chunks are input to the computation underlying inferences that listeners make about speech based on characteristics of speakers suggests that these categories, as well as lexical and phonological categories, are useful for speech perception. These abstractions allow listeners to associate incoming tokens with types on a variety of levels, so that new tokens that correspond to previously experienced tokens on some levels but not others can be accurately categorized. It appears that social information about the speaker can help to determine what constitutes a correspondence between an incoming token and previously experienced tokens.

In an exemplar model of speech perception, a nearly limitless supply of information could be used to constrain the categorization of incoming tokens. In theory, listeners could have stored details about the time of day of each utterance they hear, or the color of the walls in the room they were in when they heard it. However, many of these types of information would not be helpful to the listener in his or her quest to correctly interpret speech. Ideally, the listener would have some means of evaluating the informativity of the details of the situation accompanying the speech he or she experiences, and would be biased to use informative details to constrain their categorizations of incoming tokens, while ignoring uninformative details. A Bayesian model of language comprehension (such as that described in Norris and McQueen, 2008) builds this useful bias into the evaluation of the listener's hypothesis about what the speaker is saying, and predicts the results of Experiments 2 and 3 based on data about *t/d* deletion in production. This Bayesian approach to speech perception will be discussed in detail in the next chapter.

Chapter 7

General Discussion and Conclusion

In the first section of this dissertation, Experiment 1 showed that listeners keep track of relationships between social information and linguistic variation. This response to sociolinguistic variation is necessary for it to be a meaningful activity, and listeners must do this in order to use characteristics of speech to make inferences about the social characteristics of speakers. In the next section, Experiment 2 showed that listeners combine the information about these relationships with social information about speakers to make inferences about speech. Making use of social cues to phonetic variation is an efficient use of available information to solve the difficult problem of assigning meaning to the speech stream. In the final section, Experiment 3 showed that listeners represent their knowledge of sociolinguistic variation in a general way that can apply to novel word forms. This suggests that the inferences listeners make on the basis of sociolinguistic variation must rely on associations between social characteristics of speakers and their use of sociolinguistic variables, rather than associations between characteristics of speakers and pronunciations of specific words.

Using speech to make inferences about social characteristics of speakers (which I will call *Social Inferencing from Speech*) and using social information to make inferences about speech (which I will call *Socially Influenced Speech Perception*) can both be modeled as processes of Bayesian inference. According to Bayes' rule (see Figure 7.1), the posterior probability of a hypothesis is proportional to the prior probability of the hypothesis multiplied by the likelihood of the data observed given the hypothesis:

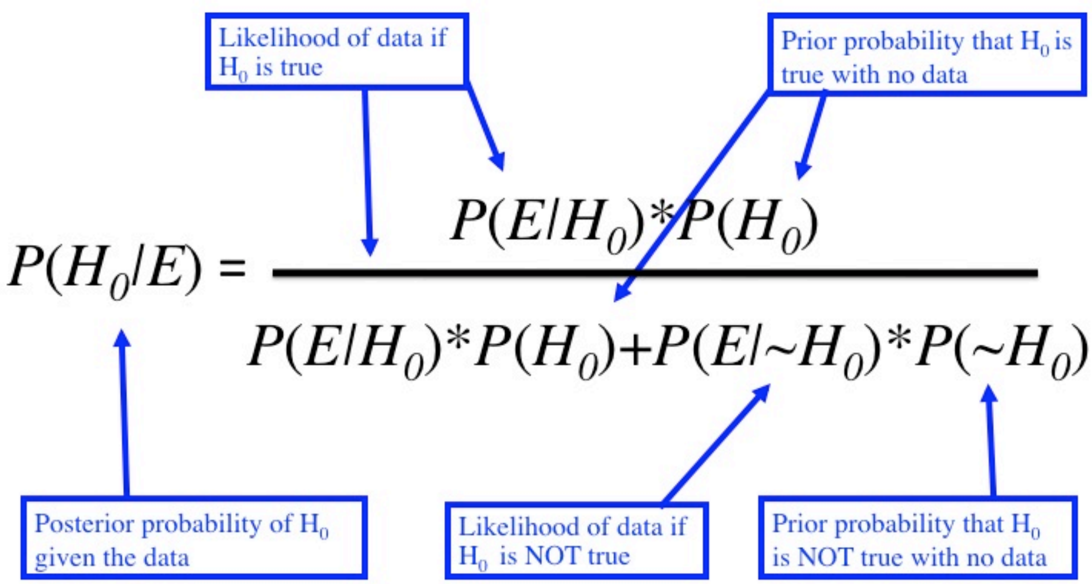

Figure 7.1. Bayes' rule. See Yudkowsky (2003) for an introduction to Bayesian reasoning and the derivation of Bayes' rule.

In the case of *Social Inferencing from Speech*, listeners use their knowledge of the relationships between social information and linguistic variation in evaluating the hypothesis that they are speaking to a person of a certain race, given that they have heard a certain pronunciation. In the case of *Socially Influenced Speech Perception*, they use this knowledge to evaluate the hypothesis that they have heard a *t/d* deletion, given the race of the speaker.

7.1 Social Inferencing from Speech

To see how this applies to the way listeners use information from the speech stream to make judgments about social characteristics of speakers (*Social Inferencing from Speech*), imagine the following situation: A listener is talking on the phone to a speaker whom they have never met before. At the beginning of the phone conversation, the listener does not know the race of the speaker, and is trying (consciously or unconsciously) to evaluate what race the speaker likely belongs to. For the sake of simplicity, I will assume that the only two possibilities being considered are that the speaker is Black or White (although this is not the situation in most real-life encounters). Then, the speaker utters the sentence *The mast probably lasted through the storm*, pronouncing the word *mast* as [mæs]. The results of Experiment 1 suggest that hearing this deleted *t* should influence the listener's estimate of how likely the speaker is to be Black. Bayes' law can provide a model of the process by which the listener uses the information about how the speaker pronounced the word *mast* (that is, the fact that the speaker deleted the final consonant) in combination with prior experience with *t/d* deletion to inform his or her beliefs about the speaker's race.

In the above example, the hypothesis that is being evaluated (H_0) is that the speaker is Black. The alternative to this, ($\sim H_0$), is in this simplified world the possibility that the speaker is White. The purpose of the model is to see how listeners can use information from the speech stream to update their prior beliefs about the speaker's race. The information from the speech stream is represented in the equation as the evidence (E); that is, the fact that the speaker deleted a *t*.

Using these values for H_0, $\sim H_0$, and E, it's possible to interpret the components of the equation in terms of the telephone call example. The term in the left side of the equation is the posterior probability of the hypothesis, $P(H_0|E)$ – that is, how likely the listener thinks the speaker is to be Black, given the fact that they have deleted a *t*. This term is what the equation is being solved for – it represents the listener's updated beliefs about the speaker's race.

The second term, $P(E|H_0)$, called the likelihood term, is the conditional probability of observing the evidence given the hypothesis. This term represents the listener's estimate of the likelihood that the listener would have heard a deleted *t*, assuming that the speaker was Black. I have specified that this term is an estimate on the part of the listener because in order to solve for the posterior probability (in other words, in order to do Bayesian inference), the listener needs to assign a value to this term.

Fortunately, it is easy for the listener to estimate the likelihood that a given type of speaker would delete a *t* by relying on past experience. For the telephone call example, I will simply choose a reasonable round number and say that the listener has heard Black speakers delete a final *t* 80% of the time that there is an opportunity to do so (leaving 20% of these cases undeleted)[9]. In this scenario, based on their experience, the listener should estimate the likelihood of a Black speaker producing a deleted token at about .8. This term is where the listener's knowledge of the rates at which Black speakers produce deleted tokens surfaces in the social inferencing process. The likelihood term is contingent on the hypothesis, so when the hypothesis concerns the speaker's race or ethnicity, then the likelihood term will be specific to speakers of a certain race or ethnicity. Importantly, this term would be different if the hypothesis were that the speaker is White.

The last term, $P(H_0)$, is the prior probability of the hypothesis given no data. In this example case, this term represents how likely the listener thought it was that the speaker was Black before getting any information about their use of *t/d* deletion. In real life, if the listener has no information at all about the race of the speaker, their

---

[9] While this figure does not reflect the exact usage of any group of Black speakers from a production study, it is not inconsistent with the possible usage of college-aged Black males according to the results of several community studies (see Chapter 2). Moreover, choosing a specific figure from a production study would not be any more accurate than this estimate, because the speech collected in these studies does not necessarily correspond to the speech heard by the participants in these experiments. In addition, the group selected for analysis in the production study does not necessarily correspond to the group over which listeners would be computing their estimates. It is likely that the estimates are computed based on the speech situation, reflecting the salience of social characteristics in the situation. If this is the case, then no one figure could be determined that listeners should use for this estimate in all situations.

estimate will probably reflect the overall likelihood in the population that a person will be Black, which in many parts of America tends to be less than 50%. People's estimates of this probability may also be influenced by markedness, leading them to assume the unmarked to an even greater degree than their experience warrants. However, in this example, I will start with the unlikely but simple scenario of a totally unbiased listener, who assigns a probability of 0.5 to each of the two possibilities. In this example, then, the product of the likelihood and the prior is 0.4 (0.8 x 0.5= 0.4).

The product of the likelihood and the prior is then divided by a normalizing constant, which is the sum of the product of the likelihood (conditional probability), and the prior probability that the hypothesis is true, $P(E|H_0)*P(H_0)$, and the product of the conditional probability and the prior probability that the hypothesis is *not* true $P(E|{\sim}H_0)*P({\sim}H_0)$. This part of the equation captures the intuition that if a speaker is not Black, they must belong to some other racial group (and in the simplified world of this example, the alternative is that they are White). If the prior probability that the speaker is Black is 0.5, then the prior probability that this is not true and that the speaker is White, ${\sim}H_0$, is also 0.5.

The conditional probability that the speaker is White given that they have deleted a final *t,* $P(E|{\sim}H_0)$, must also be estimated from experience. For the sake of the example I will suppose that listeners have heard White speakers produce a deleted final consonant 60% of the time that they have had an opportunity to delete – a smaller proportion of the time than their Black counterparts. The empirical reason to believe that these estimates are different is the results of Experiment 1, which suggest that listeners do associate *t/d* deletion more with Black speakers. It is important that this number be different from the 80% estimated for the conditional probability of H_0, because if listeners' beliefs about how often Black and White speakers engage in *t/d* deletion did not differ, then hearing a deleted final consonant should not influence their estimates of speakers' likelihoods of being Black or White.

It is worth noting that in this example, White speakers also delete final *t/d* more than 50% of the time. This situation illustrates why the dividend of the equation is not enough by itself to account for Bayesian inference. Without the normalizing constant, hearing a deleted final consonant would increase the likelihood of the speaker's being Black *and* the likelihood of the speaker's being White. Of course, this is not possible because it is not possible for a single speaker to be both Black and White[10]. The normalizing constant ensures that this paradox does not occur: it is not the absolute rate at which each type of speaker deletes but the *difference* between the rates at which different types of speakers delete that makes *t/d* deletion informative about the race of the speaker.

The ratio of the likelihoods of $P(E|H_0)$ and $P(E|{\sim}H_0)$ is a measure of the 'informativity' of the evidence (E); the higher the ratio of these two estimates, the more informative the evidence is about the hypothesis. In an efficient system, the informativity of a contextual factor about a linguistic factor would influence whether details about the contextual factor are stored along with tokens of linguistic material.

---

[10] A speaker who identifies with both races would not fit into the simplified binary categorization used in this model; the model would have to include many more categories in order to capture the many possibilities for racial identification in the U.S.

In order to evaluate the informativity of any factor with respect to another, listeners would have to store details about it. This may seem like a Catch-22: according to this logic, the listener needs to store details about a factor to determine whether or not to store details about it. However, it is sensible in a dynamic system that can adjust its sensitivity to aspects of context on the basis of continuously updated estimates of their informativity.

Using these estimates of the parameters in the model, it's possible to see how a listener's belief that Black speakers delete more final consonants (which is reflected in the different likelihood terms for Black and White speakers) could cause them to revise their estimate of a speaker's likelihood of being Black based on whether or not they produce deleted *t*. Their original estimate of this likelihood was 0.5. Substituting the estimates of the likelihood and prior terms for H_0 and $\sim H_0$ into the equation yields the result that the listener's new estimate of the speaker's likelihood of being Black is 0.57 (see Figure 7.2). The information that the speaker produced a deleted *t* caused the listener to become more confident that the speaker is Black.

Of course, each individual word that listeners hear a speaker say may not influence their beliefs about the speaker's race very much. Because *t/d* deletion is used by almost all speakers, but at different rates, production of *t/d* deletion is measured in rates over larger samples of speech, rather than individual instances. In a model of language understanding as Bayesian inference, this process of revising the listener's estimate of the likelihood that a speaker is of a certain race (that is, *Social Inferencing from Speech*) occurs iteratively, such that although it is based on individual instances, the listener's beliefs about the speaker's race are influenced by their overall *t/d* deletion rate.

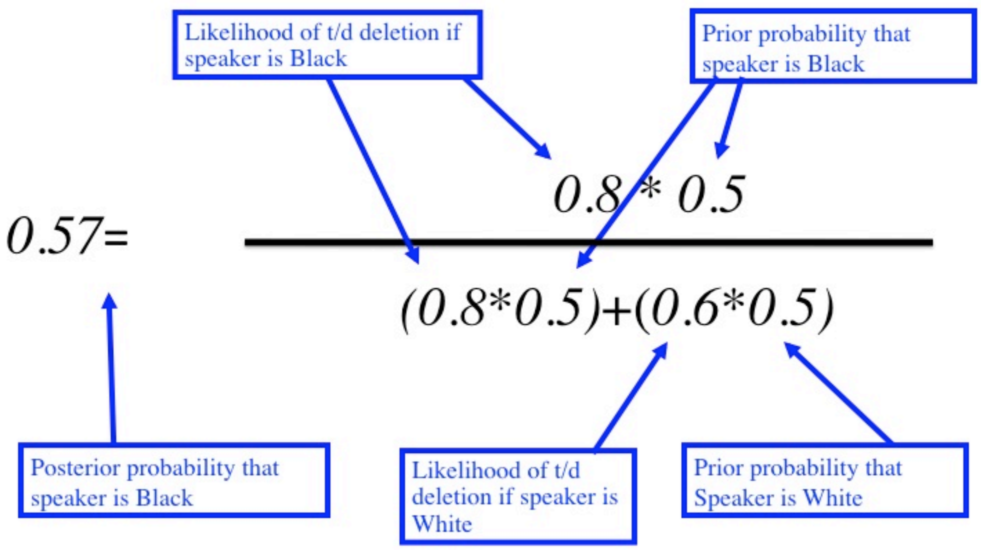

Figure 7.2. Using an estimate of 80% for a Black speaker's chance of deleting a t, *60% for a White speaker's chance of deleting a* t, *and 50% for the speaker's prior chance of being Black and their prior chance of being White, yields the result that the new chance of the speaker's being Black given that he/she deleted a* t *is 57%, higher than the prior estimate of 50%.*

7.2 Socially Influenced Speech Perception

Very similar logic applies to the complementary problem of using information about race to aid in speech processing (*Socially Influenced Speech Perception*). In a model of language understanding as Bayesian inference, determining what word has been uttered is equivalent to assigning a probability to an interpretation of the speech stream. In the case of *t/d* deletion, the hypothesis the listener must be evaluating (H_0) is that although there was no [t] in the speech signal, he or she has heard a word with an underlying /t/ at the end of a consonant cluster – for example, the word *mast*. As above, imagine that a listener and a speaker are conversing, but in this case, the conversation takes place in person, and the listener is aware that the speaker is Black (this is the evidence, E). The speaker utters the phrase *The [mas] probably lasted*, and the listener must decide whether the speaker has uttered a deleted token of the word *mast* (H_0) or a token of the word *mass*, which would not require imputing any deletion ($\sim H_0$). Thus, deciding whether the speaker intended *mast* or *mass* is equivalent to deciding whether or not the speaker has uttered a deletion. How can the listener use information about the speaker's race to assign a probability to an interpretation of the speech stream?

In this example, the posterior probability is the probability that speaker produced a deleted token of the word *mast*, given that the speaker is Black. The likelihood term represents the listener's estimate of the likelihood that the speaker would be Black, given that they have uttered a deleted token of *mast*. This estimate can be derived from the estimates from the previous example, assuming that Black and White speakers produce approximately the same number of potential final consonant clusters[11]. If Black speakers delete 80 out of 100 final consonants, and White speakers delete 60 out of 100 final consonants, then there are 140 total deleted final consonants, 80 of which

[11] Although this has not been empirically verified, it is difficult to think of a reason why Black and White speakers should differ in this respect.

were produced by Black speakers. This means that the likelihood that a deleted *t* was produced by a Black speaker is 80/140, or 57%.

It is not a coincidence that the estimate of the likelihood term in this example is the same as the posterior probability listeners assigned to the speaker's being Black in the previous example; this is simply a different way of stating the same question: given that they have produced a deleted token, what are the chances that a speaker is Black? The posterior probability of the hypothesis and the likelihood term are symmetrical, and they play opposite roles in *Social Inferencing from Speech* and *Socially Influenced Speech Perception* because these two processes are likewise symmetrical. In the previous example, the likelihood term depended on the race of the speaker because the hypothesis directly concerned the race of the speaker. In this example, the likelihood term depends on the race of the speaker because the evidence consists of the race of the speaker.

In the current example, the prior probability of the hypothesis given no data is the likelihood that the speaker would delete a *t* independent of their race. In the imaginary world I have described in which people have a 50% chance of being Black and a 50% chance of being White, and Black speakers produce deleted tokens 80% of the time and White speakers produce deleted tokens 60% of the time, this likelihood will be 0.7. So, the product of the likelihood and the prior in this example is 0.4.

The normalizing term in the divisor, as above, is the sum of the product of the likelihood (conditional probability), and the prior probability that the hypothesis is true, $P(E|H_0)*P(H_0)$, and the product of the conditional probability and the prior probability that the hypothesis is *not* true $P(E|\sim H_0)*P(\sim H_0)$. If the prior probability that the observed token was a result of a deletion is 0.7, then the prior probability that this is not true and that the token was not the result of a deletion is 0.3.

The conditional probability that the speaker is Black given that the token is *not* the result of a deletion, $P(E|\sim H_0)$, can also be derived from the previous estimates, as above. If Black speakers uttered 20 non-deleted tokens out of their 100 tokens, and White speakers uttered 40 non-deleted tokens out of their 100 tokens, then Black speakers uttered 20 out of 60 non-deleted tokens, and the likelihood of a speaker being Black given that they have uttered a non-deleted token is .33. Using all these estimates for the terms in the equation, the posterior probability that a word uttered by a Black speaker has had a final consonant deleted is 0.8 (see Figure 7.3) – this corresponds to the original 80% chance we gave a Black speaker of uttering a deleted token in the first example.

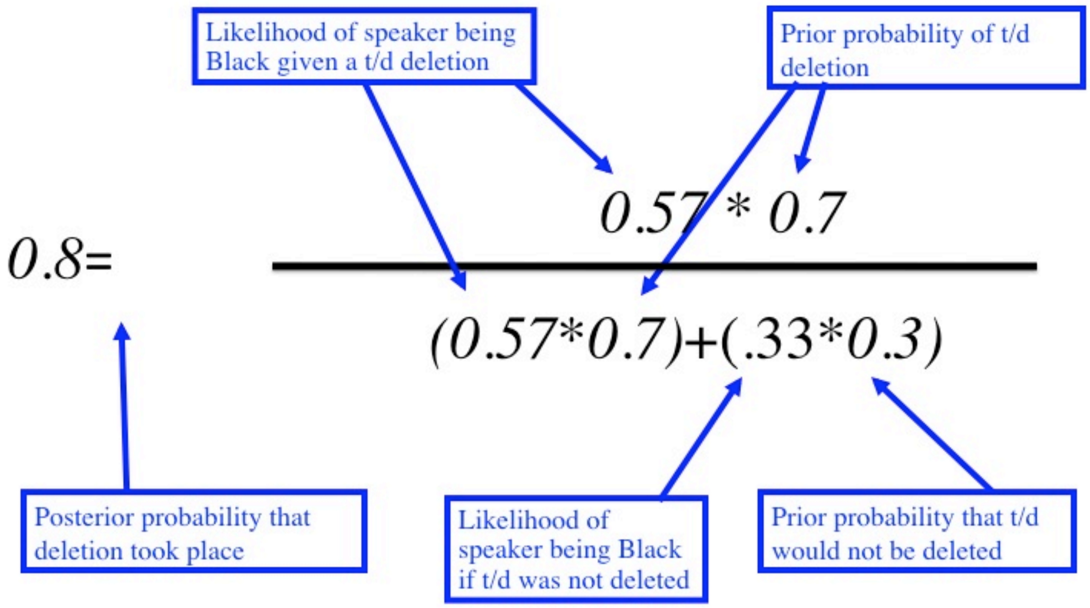

Figure 7.3. Using an estimate of .57 for the chance of a deleted t being uttered by a black speaker, .33 for the chance of a non-deleted t being uttered by a Black speaker, .7 for the token's prior chance of being deleted and .3 for the token's prior chance of being non-deleted, yields the result that the probability of a deletion having taken place given that the speaker is Black is 0.8, higher than the race-neutral estimate of 0.7.

Because of the symmetrical relationship between the process of *Social Inferencing from Speech* and *Socially Influenced Speech Perception*, the Bayesian model predicts the results of Experiment 2 based on the results of Experiment 1. Experiment 1 demonstrated that listeners attribute deleted tokens more often to Black speakers than to White speakers. This result indicates that listeners estimate a higher likelihood of the speaker being Black for deleted tokens than non-deleted tokens. The Bayesian model predicts that if this likelihood is higher, the posterior probability of interpreting the speech stream as resulting from a *t/d* deletion should also be higher.

It is impossible to directly query the probability listeners have assigned to the *mast* interpretation, which is what the Bayesian model makes a prediction about. However, there are behavioral correlates of this higher probability that can be measured. In Experiment 2, I assessed the probabilities listeners assigned to the *mast* interpretation by measuring reaction times to the sentence continuation that fit better with that word. Slower reaction times indicate lower probabilities assigned to the hypothesis suggested by the sentence continuations. The model predicts cases of Black speakers producing [mæs] when they mean *mast* (that is, Black speakers producing deleted tokens) to be high probability, and these cases do elicit faster reaction times than cases of White speakers producing [mæs] when they mean *mast*, which the model predicts to be low probability.

The Bayesian model as I have described it takes as input estimates of the relationships between race and the variable *t/d* deletion. However, it would also be possible to describe a model of the process by which social information influences speech perception on the level of individual words. Instead of one association between race and a sociolinguistic variable, the model would take as its input the relationship between race and the pronunciation of each word. I have not described such a model in detail because the results of Experiment 3 suggest that this is not the way listeners store associations between race and *t/d* deletion. However, a Bayesian model of speech perception is not in principle incompatible with word-specific knowledge of sociolinguistic variation. It only requires that listeners have associations between social information and linguistic behavior on some level, and that this information is accessible during speech perception.

As discussed in Chapter 6, the way in which behavioral responses in Experiment 3a correspond to probabilities assigned to different interpretations of the speech stream is much less straightforward than in Experiment 2a. However, if the mechanisms posited in Chapter 6 are in fact at work in the perception of spoken words and the reading of written words, then the results of Experiment 3a are also consistent with the predictions of the Bayesian model. The results of Experiment 3b are equally consistent with these predictions, provided that listeners have associations between social characteristics and properties of consonant clusters, rather than specific words.

I have proposed that using social information can help listeners interpret the speech stream, and that this might provide a functional motivation for listeners to use social information in speech perception. However, in the experiments presented in this dissertation, the use of information about the race of the speaker sometimes caused listeners to impute a *t* that was never actually present in the speech signal. This suggests that social information is not inherently helpful – being influenced by information about speaker race could potentially be detrimental to the listener. This raises the question, why would we have a speech perception system in which social information sometimes causes you to hear things that weren't there?

The answer lies in the difference between the conditions that exist in the laboratory and those that exist in the natural world. While the human speech perception system may not be perfectly adapted for the tasks listeners performed in these experiments, the lab differs crucially from real life in that social information and linguistic behavior were varied independently from one another in these experiments. In the real world, social characteristics and phonetic realizations tend to covary, which is exactly how listeners develop these different estimates of the likelihood terms in the first place. It is still possible for social information to steer a listener in the wrong direction in the real world; for example, according to the Bayesian model with the estimates I have provided for its parameters, 1/3 of the time that a listener heard a non-deleted token, it would incorrectly lead them to increase their estimate of the speaker's likelihood of being White[12]. However, as long as social factors and linguistic behavior are statistically correlated, the social information is helpful, on average.

The models I have described in this chapter are simplified in many ways relative to the situation in the real world. One of the ways in which the real world may differ from these imaginary situations is in the degree of confidence listeners might have about the evidence that is the input to these inferences. In both examples, I specified that listeners use information about one domain (race or *t/d* deletion) to make inferences about the other. In these examples, listeners have perfect information about one domain, and imperfect information about the other. However, in real life, listeners rarely, if ever, have perfect information about anything. In many situations, listeners may be using their beliefs about a speaker's use of *t/d* deletion to make inferences

[12] This is an unrealistically dire situation for the listener, because the model takes into account only one source of information, whereas listeners are presumably making inferences over many different types of evidence at once, which tends to mitigate the effects of any one piece of information in either direction.

about their race *and* using their beliefs about the speaker's race to make inferences about their use of *t/d* deletion at the same time. In such a situation, the Bayesian model makes a prediction: hearing tokens of *t/d* deletion early in an encounter should make later, ambiguous tokens more likely to be interpreted as deletions.

The prediction that earlier tokens should influence the perception of later tokens follows from the interaction of the two symmetrical processes described above. When the listener hears a token and classifies it as deleted, this increases the likelihood they assign to the speaker being Black (by *Social Inferencing from Speech*). Because the output of this process is the input to the process of *Socially Influenced Speech Perception*, this in turn makes all following tokens more likely to be interpreted as deleted tokens. Thus, when the listener does not have perfect information about either domain, the act of classifying tokens and categorizing the speaker can change the way the listener classifies tokens and categorizes the speaker in the future – an effect sometimes referred to as *perceptual learning* (Goldstone, 1998).

Perceptual learning has been observed in speech perception using several experimental paradigms; listeners are able to adapt to the speech characteristics of individual speakers (Kraljic and Samuel, 2007), foreign accents (Clarke, 2003), and even dialects (Kraljic, Brennan, and Samuel, 2008). Perceptual learning effects require listeners to be sensitive to details of the perceptual signal, as exemplar models of speech perception predict they should be. However, perceptual learning does not depend on the detailed episodic traces that underlie traditional exemplar models; perceptual learning requires only that the listener's system of perceptual categories can be influenced by these details as they are perceived. Thus, perceptual learning effects are consistent with models in which categories are emergent from perceptual experiences, whether or not these categories are constituted by detailed episodic traces of those experiences. Whether listeners engage in perceptual learning of individual sociolinguistic variables, as predicted by the Bayesian model, is a topic for further investigation.

7.3 Conclusion

The big question motivating the research reported in this dissertation concerns the nature of sociolinguistic knowledge. The question can be framed from either a sociolinguistic or a psycholinguistic point of view, but in either case, the answer requires insights and methodologies from both fields. Likewise, both fields can benefit from a better understanding of what speaker and listeners know about sociolinguistic variation, and how they use this knowledge.

The foundation of the study of linguistic variation has been documenting the factors, both linguistic and non-linguistic, that influence the way speakers produce language. Although there are functional motivations for many of the linguistic factors that constrain variability in language production, due to features of the vocal tract and articulators, among other things, there are rarely, if ever, functional motivations for sociolinguistic variation. The fact that social characteristics of speakers influence their use of linguistic variables suggests a communicative role for socially motivated variation; that is, this type of variation, rather than being epiphenomenal, is meaningful. The future of variationist sociolinguistics will involve acquiring a fuller understanding of the process by which speakers and listeners negotiate meaning; understanding this process requires modeling the listener's knowledge of sociolinguistic variation and investigating the ways in which this knowledge is used as listeners understand language.

The experiments reported in the previous chapters join a growing collection of recent efforts to expand our knowledge of what listeners know about sociolinguistic variation and how they use the information they have. Knowing that listeners associate individual linguistic variables with the social groups that use these variables provides another link in the chain of social meaning. Knowing that listeners consider social characteristics valid and useful predictors of linguistic variation indicates that listeners have expectations about the way speakers will use sociolinguistic variables. These expectations provide a backdrop for all linguistic interactions, such that speakers can then meet them to draw on existing relationships between identity and linguistic variation, or violate them to create new ones. The listener's expectations about the speaker's behavior can play an important role in defining the space of interpretable moves that the speaker can make, and thus in the communication of social meaning in general. The fact that these expectations seem to extend to novel forms suggests that listeners do not simply expect speakers to reproduce things they have already produced; rather, they expect speakers to do things that are consistent with their social positioning. These things are defined relative to the space of possible realizations of a type, rather than the space of past realizations. This highlights the importance of the envelope of variation not just for the researcher but for the listener, who relies on the space of possible realizations of a variable to interpret speech as socially meaningful, and for the speaker, who relies on the listener's interpretations for successful communication.

The importance of the envelope of variation is reflected in the Bayesian model of *Social Inferencing from Speech*, described in this chapter. The terms of the equation make explicit reference to the envelope of variation, because the value of the term $\sim H_0$ can only be estimated once a set of alternatives to H_0 has been defined. Whether or not a Bayesian model is the right way of describing the inferences listeners are making, one of the basic insights from this model must apply: Linguistic variation is only informative about a social characteristic if it is associated with one group more than with another, and social information is only informative about linguistic variation if it is associated with one variant more than another. Determining whether a social characteristic like race is a good predictor of a variant like deleted *t/d* requires considering all the cases where *t/d* is not deleted. Although defining the envelope of variation is problematic in some cases for the researcher, listeners must be doing this implicitly in order to use their knowledge of sociolinguistic variation in the ways that the experiments presented in this dissertation demonstrate that they do. Of course, the way listeners define the envelope may differ depending on their prior experiences or even their own production, and it may not be the same as the way sociolinguists define it.

There is still much to discover about how listeners acquire sociolinguistic knowledge and how they use it to make inferences about speakers and speech. How do listeners' experiences with a group or a variable influence their use of linguistic cues to make social inferences, and vice versa? Are listeners more able to interpret sociolinguistic variants that they themselves use? Do different types of listeners interpret the same speech differently? Do speakers tailor their use of sociolinguistic variables to match the knowledge they believe listeners have? One common characteristic of many of the unanswered questions regarding listeners is that they are difficult to answer by observing spontaneous interactions or even eliciting speech in an interview setting. However, many of them may be amenable to experimental investigation, and expanding the nascent subfield of experimental sociolinguistics may provide a way to keep up with the expanding space of questions sociolinguists are framing about the transmission of social meaning.

It is uncontentious that sociolinguistic knowledge is important to anyone concerned with the transmission of social meaning, but the experiments reported here demonstrate for the first time that it is also important to anyone concerned with the types of knowledge people rely on to produce and understand language. Because one of the primary concerns of psycholinguistics is how listeners comprehend sentences, there has been a substantial amount of work in this field on the way linguistic and non-linguistic information is integrated in language comprehension. Listeners have been shown to use both world knowledge gathered from prior experiences and information specific to the speech situation in comprehension. In most circumstances, listeners must combine static knowledge (such as of affordances[13] or other permanent properties of objects) with dynamic properties of the situation (such as what objects are currently available to be talked about) in order to make inferences about speech (such as what object is being referred to). For example, if a speaker says *The boy ate the…*, listeners look at edible objects like cakes more than they look at inedible objects like trucks (Altmann and Kamide, 1999). Knowing that one can only eat certain things, and knowing which things are edible, is world knowledge. The fact that there is a cake and a truck in the scene and the speaker is likely to be talking about one of them is a property of the speech situation.

In the case of sociolinguistic variation, both static and dynamic information are available: listeners have static knowledge of how linguistic variation maps onto the social landscape[14], and they also have access to dynamic properties of the social situation, such as the race or age or gender of the speaker. Thus, the nature of sociolinguistic variation makes social characteristics of speakers informative about language in much the same way that other aspects of the context are.

Despite these similarities, sociolinguistic information is unlike other types of information that are used in language comprehension in a few ways. First, much of the

---

[13] Object affordances are characteristics of an item that reflect the possible actions in which it can participate; for example, Altmann and Kamide (1999) take advantage of the fact that cakes afford eating and trucks do not.

[14] This knowledge is actually dynamic when viewed over time, perhaps more so than knowledge about things like object affordances, because relationships between social categories and linguistic variation are changed through novel uses of linguistic variation by speakers. However, it is parallel to other types of static knowledge in the sense that it is gathered from past experiences and is not dependent on the speech situation.

sociolinguistic information listeners have access to is probabilistic, because in the case of many variables, most speakers produce both variants, but at different rates. This is unlike information about things like object affordances (Tucker and Ellis, 1998), which are informative because they are either strictly consistent or inconsistent with possible interpretations of the speech being produced. Second, although sociolinguistic information is thus more like other probabilistic information that influences language comprehension, it differs from many of these types of information because there is usually a somewhat arbitrary relationship between a linguistic form and its social meaning. This is unlike, for example, the case of affective information (Nygaard and Lunders, 2002), in which a happy tone of voice leads listeners to choose a happy word meaning, or the case of articulatory information (McGurk and MacDonald, 1976), in which the speaker’s mouth shape constrains the sounds they could plausibly be making due to functional constraints, or the case of referential context (Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy, 1995; Trueswell, Sekerina, Hill, and Logrip, 1999) in which the objects present in the scene influence listeners’ syntactic interpretations because they provide different motivations for the speaker to make certain properties of the scene explicit. In all of these cases, the relationship between the information and the inferences that the listeners make could not plausibly be reversed; in general, hearing a word in a happy tone of voice would not lead listeners to choose an unhappy interpretation of that word. By contrast, in the case of sociolinguistic information, the relationships between social characteristics and linguistic behavior could sometimes be reversed, and in fact the same or similar variables have different or even opposite meanings in different populations (such as rhoticity among British vs. American English speakers). This arbitrariness is interesting, because it suggests that listeners are dependent on their experience for their static knowledge of sociolinguistic variation, due to the fact that relationships between social characteristics and speech are not inferable from other knowledge.

Given that listeners are capable of learning arbitrary associations between linguistic variation and social characteristics of speakers and then using this information to understand language, it seems sensible to ask whether the other knowledge they draw on in language comprehension is also learned via storing correlations in experience, or whether some of these types of knowledge are derived from understanding of their functional motivations. The strategy of storing correlations in experience and making inferences based on these learned associations certainly seems to be one that listeners have available to them for learning about language use, and it is one that underlies many of our most basic mental activities outside the realm of language. This suggests that many of the capacities listeners have developed for learning from and interacting in the world in general are brought to bear in language understanding.

Stored information about correlations in experience is all that is needed to perform Bayesian reasoning, and performing Bayesian reasoning is an efficient and rational way to make inferences about present and future experiences. Conceiving of the use of social knowledge in language comprehension as a process of Bayesian inference situates sociolinguistic perception as a special case of a general cognitive capacity for evaluating and using statistical evidence to understand and interact with the world.

The relationship between at least this aspect of sociolinguistics and domain-general cognitive capacities suggests an affinity between sociolinguistics and the larger enterprise of cognitive science. Sociolinguistics has been difficult to integrate with formal linguistics in some ways because of its empirical basis and its focus on the relation between language and other aspects of the world; it is just these qualities that make sociolinguistics a natural component of cognitive science, broadly construed, to which it can make valuable contributions and from which it can gain valuable insights.

Appendix

Experimental Materials

Experiments 1a, 1b, and 2b made use of 4 pictures of Black males:

And 4 pictures of White males:

Questionnaire A contained 24 potentially *t/d* deleted sentences (in their normal orthography):

1 The mast probably lasted through the storm.

2 The mist predicted by the weatherman surprised me.

3 The end probably looks like it needs trimming.

4 The least Paul made us run was three miles a day.

5 Totally flushed with shame, she ran out of the room crying.

6 The rapt lovers stared into each others' eyes.

7 He always prized the medals he won in high school.

8 The large duct had a fan at the end of it.

9 The band lifted the equipment over the railing to get it down the stairs.

10 The pond that I went fishing in was full of fish.

11 The fund never has to report any of its losses.

12 The huge find caused archaeologists all over the world to get excited.

13 The board charged my neighbor a fine for noise disturbance.

14 The silver that was mined was worth less than the recycled antique silver.

15 The first ward she wanted to visit was the maternity ward.

16 The enormous wind surprised the hikers on the cliff.

17 The mend fixed everything; Sally's shirt no longer had holes.

18 The kind of chest we usually look for in a model is smooth and muscled.

19 The graft showed signs of improvement after it had been healing for a week.

20 My first guest surprised me by bringing flowers.

21 The deft performer delighted the crowd with sleight of hand and other magic tricks.

22 The tact he had was amazing; he was polite to even the meanest people.

23 He tried to tempt me, but I no longer craved the chocolate.

24 The past should be behind us, there's no point crying over spilled milk.

And 24 matching non-*t*-word sentences:

1 The mass probably lasted an hour on Sunday.

2 The miss predicted by the announcer surprised me.

3 The “N” probably looks like an “M” in the diagram.

4 The lease Paul made us sign was three pages long.

5 Totally flush with cash, she went to the casino to celebrate.

6 The rap lovers went to a concert every day.

7 He always pries the paint can open with a crowbar.

8 The large duck had a huge bill and white feathers.

9 The ban lifted by the government was against smoking in public.

10 The pawn that I lost first was taken by the king.

11 The fun never has to end at Disney World!

12 The huge fine caused the driver of the blue sedan to get angry.

13 The boar charged the hunter and knocked him down.

14 The silver that was mine was stolen several years ago.

15 The first war she wanted to learn about was the Vietnam war.

16 The enormous win surprised the normally unsuccessful gambler.

17 The men fixed everything; they even fixed the faucet.

18 The kind of chess we usually play is a clever and fast-paced game.

19 The graph showed signs of improvement for the company's finances.

20 My first guess surprised the game show host.

21 The deaf performer delighted the crowd with sign language interpretations of opera arias.

22 The tack he had was amazing; it had the sharpest point I'd ever seen.

23 He tried to temp me, but the company wanted a permanent hire.

24 The pass should be as straight as possible, directly to the receiver.

Questionnaire B contained 24 *t/d* deleted sentences (identical to the potentially *t/d* deleted sentences except with a final *t* or *d* replaced by an apostrophe):

1 The mas' probably lasted through the storm.

2 The mis' predicted by the weatherman surprised me.

3 The en' probably looks like it needs trimming.

4 The leas' Paul made us run was three miles a day.

5 Totally flushe' with shame, she ran out of the room crying.

6 The rap' lovers stared into each others' eyes.

7 He always prize' the medals he won in high school.

8 The large duc' had a fan at the end of it.

9 The ban' lifted the equipment over the railing to get it down the stairs.

10 The pon' that I went fishing in was full of fish.

11 The fun' never has to report any of its losses.

12 The huge fin' caused archaeologists all over the world to get excited.

13 The boar' charged my neighbor a fine for noise disturbance.

14 The silver that was mine' was worth less than the recycled antique silver.

15 The first war' she wanted to visit was the maternity ward.

16 The enormous win' surprised the hikers on the cliff.

17 The men' fixed everything; Sally's shirt no longer had holes.

18 The kind of ches' we usually look for in a model is smooth and muscled.

19 The graf' showed signs of improvement after it had been healing for a week.

20 My first gues' surprised me by bringing flowers.

21 The def' performer delighted the crowd with sleight of hand and other magic tricks.

22 The tac' he had was amazing; he was polite to even the meanest people.

23 He tried to temp' me, but I no longer craved the chocolate.

24 The pas' should be behind us, there's no point crying over spilled milk.

And 24 other non-standard sentences:

1 The mailman might could deliver the mail early today.

2 The car parked near my house needs washed.

3 The fence outside my yard needs repaired, she told me.

4 I ordered a cup of cawfee at the diner last night.

5 The bucket needs filled before we can mop the floor.

6 That dog needs walked before you can go to bed.

7 I take the dawg out with me whenever I go outside.

8 My twin brother might could pass for me at school if he tried.

9 When all youse guys are gone I'm going to have to clean up the room.

10 If you just keep tawking I'll never get a word in edgewise.

11 If youse guys don't quite down, we're going to get in trouble.

12 I want the kind of dawg that doesn't bark at all.

13 What I heard from youse guys really changed my mind.

14 I might would tell you where to find the governor.

15 The trash needs emptied before the guests arrive.

16 My car's engine needs fixed before I can drive it to work tomorrow.

17 The old street needs paved before it can handle much traffic.

18 The politician might would tell us the truth if it wasn't incriminating.

19 People who need help in their life might should pray for it.

20 I told youse guys to leave me alone about that girl.

21 After youse guys took off I ran into my ex-girlfriend.

22 I drank so much cawfee the other night I stayed up for hours.

23 The professor heard youse guys talking about cheating on your assignment.

24 My mother tawks so loud she wakes up the neighbors when she whispers.

Experiment 2a made use of the same pictures that were used in Experiments 1a, 1b, and 2b, plus 8 additional female pictures:

Participants heard the beginning and then saw the end of 24 temporarily ambiguous sentences:

1 The MAST/MASS probably lasted
  a. through the storm.
  b. an hour on Sunday.

2 The MIST/MISS predicted by
  a. the weatherman surprised me.
  b. the announcer surprised me.

3 The END/N probably looks like
  a. it needs trimming.
  b. an "M" in the diagram.

4 The LEAST/LEASE Paul made us run
  a. was three miles a day.
  b. sign was three pages long.

5 Totally FLUSHED/FLUSH with
  a. shame, she ran out of the room crying.
  b. cash, she went to the casino to celebrate.

6 The RAPT/RAP lovers
  a. stared into each others' eyes.
  b. went to a concert every day.

7 He always PRIZED/PRIES the
  a. medals he won in high school.
  b. paint can open with a crowbar.

8 The large DUCT/DUCK had a
  a. fan at the end of it.
  b. huge bill and white feathers.

9 The BAND/BAN lifted

a. the equipment over the railing to get it down the stairs.
b. by the government was against smoking in public.

10 The POND/PAWN that I
a. went fishing in was full of fish.
b. lost first was taken by the king.

11 The FUND/FUN never has to
a. report any of its losses.
b. end at Disney World!

12 The huge FIND/FINE caused
a. archaeologists all over the world to get excited.
b. the driver of the blue sedan to get angry.

13 The BOARD/BOAR charged
a. my neighbor a fine for noise disturbance.
b. the hunter and knocked him down.

14 The silver that was MINED/MINE was
a. worth less than the recycled antique silver.
b. stolen several years ago.

15 The first WARD/WAR she wanted to
a. visit was the maternity ward.
b. learn about was the Vietnam war.

16 The enormous WIND/WIN surprised the

a. hikers on the cliff.

b. normally unsuccessful gambler.

17 The MEND/MEN fixed everything;

a. Sally's shirt no longer had holes.

b. they even fixed the faucet.

18 The kind of CHEST/CHESS we usually

a. look for in a model is smooth and muscled.

b. play is a clever and fast-paced game.

19 The GRAFT/GRAPH showed signs of improvement

a. after it had been healing for a week.

b. for the company's finances.

20 My first GUEST/GUESS surprised

a. me by bringing flowers.

b. the game show host.

21 The DEAF/DEFT performer delighted the crowd

a. with sleight of hand and other magic tricks.

b. with sign language interpretations of opera arias.

22 The TACT/TACK he had was amazing;

a. he was polite to even the meanest people.

b. it had the sharpest point I'd ever seen.

23 He tried to TEMPT/TEMP me,

a. but I no longer craved the chocolate.

b. but the company wanted a permanent hire.

24 The PAST/PASS should be

a. behind us, there's no point crying over spilled milk.
b. as straight as possible, directly to the receiver.

24 temporarily ambiguous fillers:

1 While Bill hunted the deer ran into the woods.

While Bill hunted the deer we made the fire.

2 They saw her duck under the fence.

They saw her duck swimming away with the ducklings.

3 The policeman asked for directions knew the way.

The policeman asked for directions to the house.

4 The old man the boats.

The old man whistles while he works.

5 The cotton clothing is usually made of grows in Mississippi.

The cotton clothing is usually breathable.

6 While Philip was washing the dishes broke.

While Philip was washing the dishes his wife was sweeping the floor.

7 The man who whistles tunes pianos.

The man who whistles tunes also sings them.

8 Fat people eat accumulates.

Fat people eat the same amount as skinny people.

9 John’s left and Mary’s right.

John’s left and Mary’s stayed.

10 John’s fast lasted three days.

John’s fast but Mary’s faster.

11 The bill was enormous, over two feet long

The bill was enormous, over two hundred dollars.

12 The track was twenty minutes long.

The track was a quarter mile long.

13 The pen was too small to write with.

The pen was too small to hold the animals.

14 The sons raise orchids.

The sun’s rays help the garden grow.

15 The meet included sprinting and long jumping.

The meat wasn’t fresh at the butcher.

16 When Fred eats food gets thrown.

When Fred eats food Jerry watches T.V.

17 She gave the child the dog bit a bandaid.

She gave the child the dog for the afternoon.

18 I convinced her children are noisy.

I convinced her children to go to bed on time.

19 She is expecting to lose the game show.

She is expecting so please don't jostle her.

20 Have the students who failed the exam take the supplementary.

Have the students who failed the exam been given detention?

21 She told me a little white lie will come back to haunt me.

She told me a little white lie at dinner.

22 The raft floated down the river sank.

The raft floated down the river and turned at the bend.

23 We painted the wall with cracks.

We painted the wall with white paint.

24 The dog that I had really loved bones.

The dog that I had really loved died last week.

And 48 unambiguous fillers:

1 When the ball goes up, the crowd holds its breath in suspense.

2 When the elevator goes down, the guard checks his surveillance monitor.

3 The high shelf has room for cookbooks.

4 The low stool matches the dining room chairs.

5 When the dough starts to rise, it'll be ready for baking.

6 If the particles floating in the beaker fall, a chemical reaction is occurring.

7 To bring up the blinds you need to pull on the strings.

8 We should bring down the screen to the child's eye-level

9 If the handle bars are raised, they'll be useless to me.

10 If the seat is lowered, the boy will be able to ride the bike.

11 The jet skyrocketed to 10,000 feet, and then leveled off.

12 The girl plummeted off of the vaulting horse.

13 When the boxes are moved up, there will be space for my clothes.

14 I want the plates in the cupboard to be moved down, and the silverware to be organized.

15 If you fix the drain in the pool, you'll see a boost in the level of the water.

16 The plane hit a patch of turbulence and experienced a drop in altitude.

17 If you see the gliders soar, try to catch them on camera.

18 When the soap bubbles sink, he laughs in excitement.

19 When the bus driver pushes the button, the wheelchair platform lifts for the passengers.

20 At track meets the flag drops to start the race.

21 If the water level is elevated, we'll know that the weather is changing.

22 If the plane's wing flaps have been depressed, you can predict that the plane is turning.

23 The climber was ascending the mountain when the snowstorm hit.

24 My grandmother was descending the stairs to try and catch it.

25 When the ball is bumped up, we'll be ready to receive it.

26 When we drove onto the shoulder the car bumped down, causing the bolt to come loose.

27 The wounded bird climbed through the foggy skies.

28 The brown squirrel tumbled through the branches.

29 When the rocket takes flight, make sure you're not standing in the open.

30 Every time Superman plunges, I get butterflies in my stomach.

31 If the window is giving you trouble, push up the glass with your hands.

32 If you want the right effect, make sure to push down the sides of the pot while the clay is still wet.

33 When the temperature goes up, many bats migrate back to their caves.

34 After the price goes down, we may reconsider our decision.

35 A high price encourages traders to act quickly.

36 A low grade compels students to work harder.

37 When my mood begins to rise, I usually listen to music.

38 If the sounds of the girls' voices fall, we can figure out where they are.

39 If you bring up the volume of the microphone, your speech will be intelligible.

40 Let's bring down the pace to accommodate the majority of the workers.

41 If the army's defenses are suddenly raised, invaders will become suspicious.

42 If our expectations are lowered, the children's progress will change.

43 The number of voters skyrocketed this election.

44 The amount of funding plummeted for our new project.

45 If your position on the list gets moved up, we'll give you a call.

46 When the team's rank moved down, they discussed future seasons.

47 The tennis player expected a boost in performance after she changed her diet.

48 The store manager expected a drop in revenue after she fired several employees.

Experiments 3a and 3b made use of the same pictures that were used in Experiment 2a, plus one additional male picture:

Experiments 3a and 3b used the 24 spoken sentence beginnings from Experiment 2a, with the *t*- and non-*t*-words serving as written prompts, as well as 24 spoken sentence beginnings that were identical except that the target words were replaced with the nonce words below (which were also used as written prompts):

1 The KROND/KRON lifted…

2 The BIST/BISS charged…

3 The kind of DEAST/DEASE we usually…

4 The REFT/REFF performer delighted the crowd…

5 The SNICT/SNICK had a huge…

6 The huge NUCT/NUCK caused…

7 Totally HOSSED/HOSS with…

8 The STIPT/STIP never has…

9 The SKEFT/SKEFF showed improvement…

10 My first FREAST/FREASE surprised…

11 The SPAST/SPASS Paul made us…

12 The FRAST/FRASS probably lasted…

13 The SMUCT/SMUCK fixed everything…

14 The silver that was FRIPT/FRIP was…

15 The CLIST/CLISS predicted…

16 The FUPT/FUP probably looks like…

17 The BLUST/BLUSS should be…

18 The GLOND/GLON that I…

19 He always SHAST/SHASS the…

20 The FIPT/FIP lovers…

21 The SPUCT/SPUCK he had was amazing…

22 He tried to SLUPT/SLUP me…

23 The first CLORD/CLORE she wanted to…

24 The enormous SKUND/SKUN surprised…

Fillers in Experiments 3a and 3b were the beginnings of the 24 ambiguous filler items and the 48 unambiguous filler items from Experiment 2a, and an additional 16 *t/d* deleted fillers, which were pronounced with a deleted final consonant in the underlined words:

1 I think the best part of…

2 Now hand me…

3 You just lift your…

4 I saw a Land Rover…

5 Don't stand next to…

6 I took some cold medicine…

7 If you spend too much…

8 Would you lend me…

9 He got in a fist fight…

10 My dog just got run…

11 I can't bend that…

12 There was a blind dog in…

13 That old thing is…

14 His fast car was in the shop…

15 I got a gift from someone…

16 He put in a hard day's…

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73* (3), 247-264.

Baugh, J. G. (1979). *Linguistic Style-shifting in Black English*. PhD Dissertation: University of Pennsylvania.

Baugh, J. (1996). Perceptions within a variable paradigm: Black and White racial detection and identification based on speech. In E. W. Schneider, *Focus on the USA* (pp. 169-182). Philadelphia: John Benjamins.

Bayley, R. J. (1991). *Variation theory and second language learning: Linguistic and social constraints on interlanguage tense-marking*. Dissertation, Stanford University, School of Education.

Bell, A. (1984). Language Style as Audience Design. *Language in Society, 13* (2), 145-204.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113* (2), 1001-10024.

Birch, S., Pollatsek, A., & Kingston, J. (1998). The Nature of the Sound Codes Accessed by Visual Language. *Journal of Memory and Language, 38*, 70-93.

Bourdieu, P. (1977). The economics of linguistic exchanges. *Social Science Information, 16* (6), 645-668.

Bourdieu, P. (1984). *Distinction*. Cambridge, MA: Harvard University Press.

Bourdieu, P., & Passeron, J.-C. (1977). *Reproduction in education, society and culture*. London: Sage.

Bradac, J. J., & Giles, H. (1988). Lexical diversity and magnitude of convergent versus divergent style shifting: Perceptual and evaluative consequences. *Language and Communication, 8*, 213-228.

Buchstaller, I. (2004). Putting perception to the reality test: The case of go and like. . *Penn Working Papers in Linguistics: Selected Papers from NWAVE 32, 10* (2).

Campbell-Kibler, K. (2005). *Listener Perceptions of Sociolinguistic Variables: The case of (ING)*. Dissertation, Stanford University.

Campbell-Kibler, K. (2007). Accent, (ING) and the social logic of listener perceptions. *American Speech, 82* (1), 32-64.

Carpenter, J. (2005). The Invisible Community of the Lost Colony: African American English on Roanoke Island. *American Speech,* 80 227-56.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *TRENDS in Cognitive Science , 10* (7), 287-291.

Cheyne, W. M. (1970). Stereotyped reactions to speakers with Scottish and English regional accents. *British Journal of Social and Clinical Psychology , 9*, 77-79.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper Row.

Clarke, C. (2003). *Processing Time Effects of Short-term Exposure to Foreign Accented English*. Dissertation, University of Arizona.

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception and Psychophysics , 70*, 604-618.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers , 25* (2), 257-271.

Creber, C., & Giles, H. (1983). Social context and language attitudes: The role of formality-informality of the setting. *Language Sciences , 5* (155-183).

Drager, K. (2006). From Bad to Bed: the relationship between perceived age and vowel perception in New Zealand English. *Te Reo* .

Eckert, P. (1989a). *Jocks and burnouts: Social categories and identity in the high school*. New York: Teachers College Press.

Eckert, P. (1989b). The whole woman: sex and gender differences in variation. *Language Variation and Change*.

Eckert, P. (2000). *Linguistic Variation as Social Practice: The linguistic construction of identity in Belten High*. Oxford: Blackwell .

Eckert, P. (2008). Variation and the Indexical Field. *Journal of Sociolinguistics*.

Fasold, R. (1972). *Tense marking in Black English: A linguistic and social analysis*. Arlington, VA: Center for Applied Linguistics.

Fasold, R. (1990). *The Sociolinguistics of Language*. Cambridge, MA: Basil Blackwell.

Fishman, J. A., Cooper, R. L., & Ma, R. (1971). *Bilingualism in the Barrio*. Bloomington: Indiana University Press.

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.

Fowler, C. A. (1980). Coarticulation and Theories of Extrinsic Timing. *Journal of Phonetics, 8*, 113-133.

Fox, R., & Nissen, S. (2005). Sex-Related Acoustic Changes in Voiceless English Fricatives . *Journal of Speech, Language, and Hearing Research, 48*, 753-765.

Fraser, B. (1973). Some 'unexpected' reactions to various American-English dialects. In R. Shuy, & R. Fasold, *Language Attitudes: Current Trends and Prospects* (pp. 28-35). Washington, DC: Georgetown University Press.

Fraser, H. (1997). Phonology without tiers: Why the phonetic representation is not derived from the phonological representation. *Language Sciences, 19* (2), 101-137.

Garrett, P., Coupland, N., & Williams, A. (2003). *Investigating Language Attitudes: Social Meanings of Dialect, Ethnicity and Performance*. University of Wales Press.

Giles, H., & Coupland, N. (1991). *Language: Contexts and Consequences*. Open University Press.

Giles, H., Coupland, N., Henwood, K., Harriman, J., & Coupland, J. (1990). The social meaning of RP: An intergenerational perspective. In S. Ramgaran, *Studies in the Pronunciation of English: A Commemorative Volume in Honor of A. C. Gimson* (pp. 191-211). New York: Routledge.

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

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Goldstone, R. L. (1998). Perceptual Learning. *Annual Review of Psychology, 49*, 585-612.

Gregory, M., Raymond, W. D., Bell, A., Fossler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *Proceedings of CLS 35*.

Grice, H. P. (1975). Logic and Conversation. *Syntax and Semantics, 3*, 41-58.

Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In Labov, *Locating Language in Time and Space* (pp. 1-36).

Guy, G. R. (1991). Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change, 3*, 1-22.

Guy, G. R., & Boberg, C. (1997). Inherent variability and the obligatory contour principle. *Language Variation and Change,* 9(2): 149-164.

Guy, G. R., & Boyd, S. (1990). The development of a morphological class. *Language Variation and Change, 2*, 1-18.

Hale, J. (2001). A Probabilistic Earley Parser as a Psycholinguistic Model. *Proceedings of 2nd NAACL* .

Harris, J. (2003). Release the captive coda: the foot as a domain of phonetic interpretation. *Laboratory Phonology, 6*, 103-129.

Hartigan, J. (2005). Culture against Race: Reworking the Basis for Racial Analysis. *South Atlantic Quarterly*, 104(3), 2005, pp. 543-60.

Hay, J., & Drager, K. (Under Review). Stuffed Toys and Speech Perception.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: exemplar priming in speech perception. *The Linguistic Review, 23* (3), 351-379.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34* (4), 458-484.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels . *Journal of the Acoustical Society of America, 97* (5), 3099-3111.

Hinton, L. N., & Pollock, K. E. (2000). Regional Variations in the Phonological Characteristics of African American Vernacular English. *World Englishes, 19* (1), 59-71.

Hubbell, A. F. (1950). *The Pronunciation of English in New York City*. New York: Columbia University Press.

Hudson, R. (1997). Inherent variability and linguistic theory. *Cognitive Linguistics* 8: 73-108

Jacewicz, E., Fox, R. A., & Salmons, J. (2007). Vowel Space Areas Across Dialects and Gender. *Proceedings of ICPhS XVI*, 1465-1468.

Johnson, F. L., & Buttny, R. (1982). White listeners' responses to "sounding Black" and "sounding White": The effects of message content on judgments about language. *Communication Monographs, 49*, 33-49.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. Mullenix, *Talker Variability in Speech Processing* (pp. 145-166). San Diego: Academic Press.

Johnson, K. (2005). Decisions and mechanisms in exemplar-based phonology. *UC Berkeley Phonology Lab Annual Report*, 289-311.

Johnson, K., Strand, E., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*, 359-384.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108*, 1252–1263.

Killgore, W., Casasanto, D., Maldjian, J., & Detre, J. (2000). Associative memory for faces activates the left amygdala and hippocampus. *NeuroReport, 11*, 2259-2263.

Kirchner, R. (1997). Contrastiveness and faithfulness. *Phonology, 14*, 83-111.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59* (5), 1208-1221.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194-231.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56* (1), 1-15.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating Variation: Dialects, Idiolects, and Speech Processing. *Cognition, 107* (1), 54-81.

Kramer, C. (1974). Wishy-washy mommy talk. *Psychology Today*, pp. 82-85.

Kroch, A. (1978). Toward a theory of social dialect variation. *Language In Society, 7*, 17-36.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. (1972). *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1973). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1975). The quantitative study of linguistic structure. In KH Dahlstedt (ed) *The Nordic languages and modern linguistics*. Stockholm: Almqvist & Wiksell, 188-244.

Labov, W. (1989). The child as linguistic historian. *Language Variation & Change,* 1(1): 85-97.

Labov, W. (1994). *Principles of Linguistic Change: Internal Factors*. Blackwell.

Labov, W. (1996). The Organization of Dialect Diversity in North America. *Proceedings of the Fourth International Conference on Spoken Language Processing*. Philadelphia, PA.

Labov, W. (2001). *Principles of Linguistic Change: Social Factors*. Wiley-Blackwell.

Labov, W. (To appear). *Quantitative Reasoning in Linguistics*. In Trudgill, P. (ed.), Mouton De Gruyter Handbook.

Labov, W., Cohen, P., Robins, C., & Lewis, J. (1968). *A study of the non-standard English of Negro and Puerto Rican speakers in New York City*. Final report, Cooperative Research Project 3288. Vols I and II.

Labov, W., Yaeger, M., & Steiner, R. (1972). *The Quantitative study of sound change in progress*. Philadelphia: US Regional Survey.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology, 60*.

Lavandera, B. (1978). Where does the sociolinguistic variable stop? *Language in Society, 7*, 171-183.

Lesch, M., & Pollatsek, A. (1998). Evidence for the use of assembled phonology in accessing the meaning of printed words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24* (3), 573-92.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106* (3), 1126-1177.

Lim, L., & Guy, G. R. (2005). The limits of linguistic community: speech styles and variable constraint effects. . *Penn Working Papers in Linguistics, 10.2* (Papers from NWAVE 32), 157-170.

Lippi-Green, R. (1997). *English with an Accent: Language, Ideology and Discrimination in the United States*. Routledge.

Luce, P., & Lyons, E. (1998). Specificity of memory representations for spoken words. *Memory and Cognition, 26*, 708-715.

Luo, C. (1996). How is word meaning accessed in reading? Evidence from the phonologically mediated interference effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 123*, 107-128.

Macaulay, R. K. (1978). Variation and Consistency in Glaswegian English. In P. Trudgill, *Sociolinguistic Patterns in British English* (pp. 132-43).

Mallinson, C. & Wolfram, W. (2002). Dialect accommodation in a bi-ethnic mountain enclave community: More evidence on the earlier development of African American English. *Language in Society*, 31: 743- 775.

McClennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of Lexical Form. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29* (4), 539-553.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14* (6), 648-652.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73-103.

Murray, T. E., Frazer, T. C., & Simon, B. L. (1996). Need + Past Participle in American English. *American Speech, 71* (3), 255-271.

Neu, H. (1980). Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. In William Labov, ed., *Locating language in time and space*. New York: Academic Press: 37-54.

Nguyen, N., & Hawkins, S. (1999). Implications for word recognition of phonetic dependencies between syllable onsets and codas. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey, *Proceedings of the XIVth International Congress of the Phonetic Sciences*. (Vol. 1, pp. 647-650). Berkeley, CA: University of California.

Niedzielski, N. A. (1996). Acoustic analysis and language attitudes in Detroit and Windsor. *Pennsylvania Working Papers in Linguistics, 3*, 73-86.

Niedzielski, N. A. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*, 62-85.

Niedzielski, N. A. (2001). Chipping away at the production/perception interface. In D. E. Jones, *NEWER Ways of Analyzing Variation*.

Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review, 115* (2), 357-395.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204-238.

Nygaard, L. C., & Pisoni, D. B. (1995). Speech Perception: New Directions in Research and Theory. In J. L. Miller, & P. D. Eimas, *Handbook of Perception and Cognition: Speech, Language and Communication*. San Diego: Academic Press.

Nygaard, L., & Lunders, E. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory and Cognition, 30*, 583-593.

Omi, M., & Winant, H. (1986). *Racial Formation in the United States: From the 1960s to the 1980s*. NY: Routledge.

Patrick, P. L. (1999). *Urban Jamaican Creole: Variation in the mesolect*. (Varieties of English Around the World, G17.) Philadelphia & Amsterdam: John Benjamins.

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

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper, *Frequency effects and emergent grammar*. Amsterdam: John Benjamins.

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics, 34*, 516-530.

Plichta, B., & Preston, D. R. (2005). The /ay/s have it: the perception of /ay/ as a North-South stereotype in US English. (T. Kristiansen, N. Coupland, & P. Garrett, Eds.) *Acta Linguistica Hafniensia, 37*, 243-285.

Plichta, B., & Rakerd, B. (2002). Perceptions of /a/-Fronting across Two Michigan Dialects. *Presented at NWAV*. Stanford, CA.

Podesva, R. (2003). *The Role of Closure Silence and Release Bursts in the Perception of Word-Final Coronal Stops*. Stanford University.

Podesva, R. (2006). *Phonetic Detail in Sociolinguistic Variation: Its Linguistic Significance and Role in the Construction of Social Meaning*. Dissertation, Stanford University.

Podesva, R. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics, 11* (4), 478-504.

Podol, J., & Salvia, J. (1976). Effects of visibility of a prepalatal cleft on the evaluation of speech. *Cleft Palate Journal, 13*, 361-366.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18*, 10-30.

Reynolds, W. (1994). *Variation and phonological theory*. Dissertation, University of Pennsylvania.

Rickford, J. R. (1979). *Variation in a creole continuum: Quantitative and implicational approaches*. Ph.D. Dissertation, University of Pennsylvania.

Rickford, J. R. (1986). The need for new approaches to social class analysis in sociolinguistics. *Language and Communication, 6* (3), 215-221.

Rickford, J. R. (1999). *African American Vernacular English: Features, Evolution, Educational Implications*. Blackwell.

Roberts, J. (1995). *The acquisition of variable rules: t,d deletion and -ing production in preschool children*. Dissertation, University of Pennsylvania.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*, 511-531.

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of non-native English speaking teaching assistants. . *International Journal of Intercultural Relations, 14*, 337-353.

Santa Ana, O. (1992). Chicano English evidence for the exponential hypothesis: A variable rule pervades lexical phonology. *Language Variation and Change,* 4(3): 275-288.

Schwartz, M. (1968). Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America, 43* (5), 1178-1179.

Schwartz, R., Shen, W., Campbell, J., Paget, S., Vonwiller, J., Estival, D., et al. (2007, August 27). Construction of a phonotactic dialect corpus using semiautomatic annotation. *Proceedings of INTERSPEECH 2007*, 942-945.

Shuy, R., & Williams, F. (1973). Stereotyped attitudes of selected English dialect communities. In R. W. Shuy, & R. W. Fasold, *Language Attitudes: Current Trends and Prospects* (pp. 85-97). Washington, DC: Georgetown University Press.

Smitherman, G. (1986). *Talkin and Testifyin: The Language of Black America*. Wayne State University Press.

Sridhara, A. (1984). A study of language stereotype in children. *Journal of Psychological Researches , 28*, 45-51. Strand, E. (2000). *Gender Stereotype Effects in Speech Processing*. Dissertation, The Ohio State University.

Strand, E. (1999). Uncovering the roles of gender stereotypes in speech perception. (L. Milroy, & D. R. Preston, Eds.) *Journal of Language and Social Psychology, 18*.

Strange, W. (1999). Perception of vowels: Dynamic constancy. In J. M. Pickett, *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Needham Heights, MA: Allyn & Bacon.

Strongman, K. T., & Woosley, J. (1967). Stereotyped reactions to regional accents. *British Journal of Social and Clinical Psychology, 6*, 164-167.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322 – 338.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*, 1086-1100.

Taft, M. (2006). Orthographically Influenced Abstract Phonological Representation: Evidence from Non-rhotic Speakers. *Journal of Psycholinguistic Research, 35* (1), 67-78.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language In Society, 1*, 179-96.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73*, 89-134.

Tucker, M., & Ellis, R. (1998). On the Relations Between Seen Objects and Components of Potential Actions . *Journal of Experimental Psychology: Human Perception and Performance, 24* (3), 830-846.

Urban Dictionary. (2008). Retrieved from http://www.urbandictionary.com/define.php?term=dawg

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

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition, 15*, 181-198.

Van Orden, G. C. (1991). Phonologic mediation is fundamental to reading. In D. Besner, & G. Humphreys, *Basic processing in reading: Visual word recognition* (pp. 77-103). Hillsdale, NJ: Erlbaum.

Vitevich, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word recognition. *Psychological Science,* 9:325:329.

Warren, R. M. (1970). Perceptual restorations of missing speech sounds. *Science, 167*, 392-393.

Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American, 223*, 30-36.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science, 146*, 1149-1151.

White, M. J., Vandiver, B. J., Becker, M. L., Overstreet, B. G., Temple, L. E., Hagan, K. L., et al. (1998). African-American evaluations of Black English and Standard American English. *Journal of Black Psychology, 24*, 60-75.

Williams, F., Whitehead, J. L., & Miller, L. (1972). Relations between Language Attitudes and Teacher Expectancy. *American Educational Research Journal, 9* (2), 263-277.

Wolck, W. (1973). Attitudes toward Spanish and Quechua in bilingual Peru. In R. Shuy, & R. Fasold, *Language Attitudes: Current Trends and Prospects* (pp. 129-147). Washington, DC: Georgetown University Press.

Wolfram, W. (1969). *A sociolinguistic description of Detroit Negro speech.* Washington, DC: Center for Applied Linguistics.

Wolfram, W. (1969b). Sociolinguistic Implication for Educational Sequencing. In R. Fasold, & R. Shuy, *Strategies for Teaching Standard English to Black Children.* Washington, DC: Center for Applied Linguistics.

Wolfram, W., & Fasold, R. W. (1974). *The Study of Social Dialects in American English*. Englewood Cliffs, NJ: Prentice-Hall.

Wolfram, W., Hazen, K., & Tamburro, J. R. (1997). Isolation within isolation: A solitary century of African America Vernacular English. *Journal of Sociolinguistics,* 1:7-38.

Woolard, K. A., & Gahng, T.-J. (1990). Changing language policies and attitudes in autonomous Catalonia. *Language In Society, 19*, 311-330.

Yudkowsky, E. S. (2003). *An Intuitive Explanation of Bayesian Reasoning*. Retrieved from http://yudkowsky.net/bayes/bayes.html

Zhang, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language In Society, 34*, 431-466.