THE LINGUISTIC AND THE SOCIAL INTERTWINED: LINGUISTIC CONVERGENCE TOWARD SOUTHERN SPEECH

Lacey Wade

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

Meredith Tamminga, Assistant Professor of Linguistics

Graduate Group Chairperson

Eugene Buckley, Associate Professor of Linguistics

Dissertation Committee

David Embick, Professor of Linguistics

Gareth Roberts, Assistant Professor of Linguistics

THE LINGUISTIC AND THE SOCIAL INTERTWINED: LINGUISTIC CONVERGENCE TOWARD SOUTHERN SPEECH

For Oliver Baz Wade.

Acknowledgments

During the work on this dissertation, and throughout my graduate career, I have been incredibly lucky to have been surrounded by mentors who somehow manage to be simultaneously brilliant, encouraging, and kind. All the thanks in the world go to my doctoral advisor and dissertation supervisor, Meredith Tamminga, whose advice is always careful, practical, and strategic, and whose constant affirmations kept me going. Being a mother in academia is never easy, but I was lucky to have such a supportive advisor who welcomed my sometimes fussy son into meetings, and who cared as much about my mental and emotional health as my academic success. Thanks for teaching me the art of pouring out the contents of my brain onto a whiteboard (though I'm afraid wall-length whiteboards are now a luxury that I can't live without), for taking my research ideas seriously, and for helping me turn them into reality. The dissertation has also benefited greatly from the feedback of my committee members. Dave Embick was extremely helpful in talking out experimental design issues and always gave me a new angle from which to think about my research. I am also grateful to Gareth Roberts, who would hold my son during meetings so I could work, and who would let him explore the contents of his office once he was able to crawl. His careful reading and focus on precision and clarity have been much appreciated.

I credit the sociolinguists at North Carolina State University, where I completed my M.A., for really shaping who I wanted to be as a researcher and providing me with the most incredibly supportive environment that laid the foundation for my academic career. Erik Thomas, my thesis advisor, taught me what a spectrogram was—and my life hasn't been the same since. I am grateful for his earnest interest in my work and for encouraging me to pursue a Ph.D. Walt Wolfram is the kind, caring, encouraging academic I aspire to be, and I'd especially like to thank him for letting me borrow his office for a week to run participants for the pilot experiments that ended up informing the dissertation work. The patient and brilliant Robin Dodsworth taught me R, encouraged me to think critically, and told me not to be afraid of quantitative work—you were right. Thanks also to Jeff Mielke for helping to shape my research interests and for always giving spot-on feedback.

I'm also grateful to a number of other faculty members at Penn who—whether through classes, reading groups, lab meetings, or practice talks—have made me a better academic: Eugene Buckley, Jianjing Kuang, Tony Kroch, Bill Labov, Julie Legate, Rolf Noyer, Don Ringe, Gillian Sankoff, Katie Schuler, and Charles Yang. A special thanks also to Amy Forsyth, our departmental coordinator, for somehow managing to keep everyone on track to do everything they are supposed to be doing all of the time, and for the sometimes much needed distraction from work to talk about Project Runway—I dread to think of all the deadlines I would have missed without her firm reminders.

There are also a number of people that I simply couldn't have completed the dissertation (at least not in any sort of reasonable time frame) without. Thanks to my two fabulous RAs, Sadie Butcher and Leila Pearlman who spent hours upon hours editing TextGrids and running participants, and to my Southern model talker Vanessa Sims for lending me her time and her voice. Thanks also to Jérémy Zehr for building (along with Florian Schwarz) the Penn Controller just in time for me to use it to collect audio data online, and for responding so helpfully to my many emails.

I also wish to acknowledge my classmates, who made my time at Penn truly enjoyable. Thanks to the three sociolinguists who came before me, all of whom I had the pleasure of sharing an office with at various points in time. Thanks to Sabriya Fisher, Duna Gylfadottir, and Betsy Sneller for their support and encouragement, and for providing a model of sociolinguistic work to aspire to. I am grateful to have learned from each of them throughout the course of my graduate career. Thanks also to my current office mates, Wei Lai, Yosiane White, Ruaridh Purse, and Aini Li, for the discussions of statistical modeling, experimental design, and linguistic theory, and for keeping Oliver entertained when I brought him in to work. Thanks to the members of my cohort. We deemed ourselves the "smart-beginners" soon after meeting, based on a description of graduate students in some book we found: Spencer Caplan, Andrea Ceolin, Ava Creemers, Jordan Kodner, Wei Lai, Caitlin Richter, Jia Tian, and Hong Zhang. Thanks also to Luke Adamson, Faruk Akkuş, Nikita Bezrukov, Ryan Budnick, Nattanun (Pleng) Chanchaochai, Aletheia Cui, Kajsa Djärv, Amy Goodwin Davies, Ava Irani, Nari Rhee, Ollie Sayeed, Milena Šereikaitė, and Rob Wilder. I'm grateful for the study groups, the parties, and the friendship over the years.

Many thanks go to my family, for still not quite understanding exactly what it is I do, but for being proud regardless. Thanks to my parents for encouraging me to follow my interests and for constant love and support. Thanks also to my parents in-law, for showing interest in what I do and for understanding when, during this last stretch of dissertation, I spent a good portion of family visits in the corner, on my computer, working on the dissertation. And thanks to both sets of parents for the more tangible gifts of babysitting, helping us move (often across state lines and on multiple occasions), and staying with us after Oliver was born. Thanks also to my siblings for their love and support. Many thanks to my many grandparents, who have been a constant source of love and encouragement, and who never stop reminding me that they are proud of my accomplishments. I'd especially like to acknowledge my grandfather Sherman Morgan, who left us in the spring of my third year, but who would have been particularly proud of the completion of this dissertation.

I finished my dissertation during the global COVID-19 pandemic, working from home for the last 6 weeks leading up to the defense with my husband, my two-year-old, and my cats. My family have made surprisingly good office mates. Thanks to my husband Drew, for being the on-duty parent every free moment you had, for keeping me filled to the brim with tea and coffee through late nights and long weekends, for impromptu living room dance parties, for absorbing all of the dissertation-related stress I spewed out throughout the last few months, and for constant words of encouragement. I truly couldn't have done it without you. And, finally, to my son, Oliver, who joined us at the end of my third year. While taking care of a toddler (stuck at home during quarantine) didn't exactly make the dissertation writing *easier* (I can't count the number of times I found myself sitting, content, staring at this amazing little human asleep on my chest, instead of taking him to bed so I could write), he certainly made the experience more enjoyable. I am grateful to you for so many reasons. For all the stress-melting snuggles and for the entertaining moments of language acquisition (whose mantra "No like it," which later became simply "Nike it," described well our mutual feelings toward my spending hours in front of the computer instead of playing). Above all, thank you for teaching me balance, for making much-needed breaks from work a priority, and for putting everything into perspective.

ABSTRACT

THE LINGUISTIC AND THE SOCIAL INTERTWINED: LINGUISTIC CONVERGENCE TOWARD SOUTHERN SPEECH

Lacey Wade

Meredith Tamminga

The dissertation examines the relationship between social and linguistic knowledge using a series of experiments eliciting linguistic convergence to Southern speech. I draw a terminological and theoretical distinction between previously observed *input-driven convergence*, in which speakers converge toward a linguistic form directly observed in the input, and *expectation-driven convergence*, in which speakers converge toward a linguistic form they only expect but do not observe in the immediate input. Using a novel Word Naming Game paradigm designed to elicit convergence toward expected rather than observed linguistic behavior, Experiment 1 finds experimental evidence for expectation-driven convergence, which had previously only been anecdotally observed; participants converge toward glide-weakened /aɪ/, a salient feature of Southern English, which they may expect but never directly observe from a Southern-accented model talker. The existence of expectation-driven convergence suggests that accounts of convergence relying on tight perception-production links where production is derived directly and automatically from the input cannot straightforwardly explain all instances of convergence. Experiment 2 investigates the perceptual underpinnings of input- and expectation-driven convergence using an auditory lexical decision task in which participants judge glide-weakened /aɪ/ items (e.g., *bribe* produced as *brahb*) as words or non-words. I find higher word-endorsement rates for glide-weakened /aɪ/ words for participants who have recently heard a Southern-accented (compared to Midland-accented) talker, even if the Southern talker never produces the /aɪ/vowel. Individual perception and production responses toward glide-weakened /aɪ/ show little evidence for strong individual perception-production links, though findings are consistent with an interpretation where

perceptual shifts are a necessary (but not sufficient) precursor to production shifts. Finally, Experiment 3 uses a dialect-label manipulation version of the Word Naming Game and demonstrates that both top-down information about social categories and bottom-up acoustic cues independently contribute to expectation-driven shifts in production and perception. Further, reliance on these cues differs across dialect backgrounds, providing insights into the way sociolinguistic associations are formed and mentally represented. Taken together, findings support a model of cognition in which social and linguistic information are tightly linked.

Contents

List of Tables

List of Figures

Chapter 1

Background

Five decades ago, William Labov objected to the term “sociolinguistics,” on the grounds that there is no way to truly study language without taking social dimensions into account. Since then sociolinguists have continued to find that social and linguistic information are inextricably linked. Linguistic features have been shown to index various social meanings (e.g., Campbell-Kibler, 2009; Eckert, 2000; Campbell-Kibler, 2011; Mendoza-Denton, 2008). Conversely, individuals have been shown to possess implicit social knowledge that can influence the way language is perceived (e.g., Niedzielski, 1999; Hay et al., 2006a).

Discovering the mental reality underlying linguistic behavior has been a goal of linguists for at least the latter half of the last century (e.g., Chomsky, 1965). However, a focus on abstract properties and categorical phenomena means linguistic variability has often been viewed as a mere complication and discarded. Though there has been some work on the mental representations of variability in linguistic theory, such as the variable rule framework (Labov, 1969; Cedergren and Sankoff, 1974), much less attention has been paid to the representation of the social factors inherent in variability and how they interact with linguistic representations. As Thomas (2011) notes, “The result is cognitive theories that cannot account for variation adequately and variationist theories that do not address the mental structure of language. This situation has not changed since Labov made the same observation in 1975.” While there has been some improvement since Thomas’ observation in 2011, this is still a relatively new area of inquiry.

This dissertation aims to shed light on the mental representations of social factors, and how these are linked to linguistic representations, through a series of experimental investigations. I use linguistic convergence, when language users shift their speech to become more similar to another talker, as a window into the relationship between linguistic and social knowledge. In Experiment 1, I report evidence that speakers are able to converge toward a phonetic variant that they may associate with—but do not directly observe from—a model talker, suggesting that socially-cued linguistic expectations influence linguistic production. That is, speakers' pre-existing knowledge about sociolinguistic associations informs their own speech production. In Experiment 2, I suggest that this type of *expectation-driven* convergence has perceptual underpinnings, and differs from *input-driven* convergence in a number of ways. In Experiment 3, I further elucidate the cognitive mechanisms behind this behavior and propose that both socially-rooted "top-down" mechanisms and structurally-rooted "bottom-up" mechanisms may contribute to expectation-driven convergence.

1.1 Linguistic and Social Knowledge

The relationship between the linguistic and the social was recognized relatively early on. In one of the first quantitative sociolinguistic studies, Fischer (1958), examined the social factors that influence variation of IN and ING in New England schoolchildren. He found that boys produced more of the non-standard IN variant, and that this finding stemmed from gender differences in conformity to linguistic norms. He further found that "typical boys" produced more of the non-standard IN variant than "model boys." On Martha's Vineyard, Labov (1963) found that centralization of the /au/ and /aɪ/ diphthongs correlated with a desire to identify oneself as a Vineyarder. Later, Labov (1966) found that use of vocalic /r/ varied in predictable ways based on the social class of the language user, and the amount of attention paid to speech, such that higher social classes and more attention paid to speech correlated with increased rates of *r.* Rather than exemplifying "free variation" as had previously been assumed, these early studies all showed that previously unexplained linguistic variation patterned with the social attributes of the language users; this is referred to as *orderly heterogeneity* (Weinreich et al., 1968). Early studies providing evidence for orderly heterogeneity were a first step toward understanding the cognitive relationships between the social and the linguistic, showing that these two dimensions vary together in predictable ways, though much of the focus was on the actual facts of language usage across various social dimensions and not as much on the way linguistic features are socially perceived. More recently, sociolinguistics has shifted attention to the fact that, not only do the social and the linguistic correlate, but they are linked in the minds of language users.

Third Wave approaches to the study of sociolinguistics, for instance, shift the focus from static macro-social demographic categories to real-time construction of social meaning. As Eckert (2012) notes, “variables cannot be consensual markers of fixed meanings; on the contrary, their central property must be indexical mutability. This mutability is achieved in stylistic practice, as speakers make social-semiotic moves, reinterpreting variables and combining and recombining them in a continual process of bricolage” (p. 94). Demonstrating that speakers utilize varying linguistic forms to index a wide range of (ever-changing) social meanings suggests that the linguistic and the social do not just happen to covary, but that they must be linked together in a more substantial way in speakers’ awareness. After all, speakers cannot be expected to use linguistic variants agentively to index particular social meanings if they do not have knowledge about the relationship between a given linguistic form (or combination of forms) and social meanings. As Eckert (2012) puts it, “Whereas the first two waves viewed the meaning of variation as incidental fallout from social space, the third wave views it as an essential feature of language” (Eckert, 2012, p. 94).

The ability of linguistic forms to actually signal social meaning has not just been inferred based on ethnographic work, but can actually be experimentally tested. For instance, the primary tool for isolating particular social meanings attached to speech styles or variants is the Matched Guise task, for which participants rate various “guises” that differ in the linguistic dimension of interest on various social attributes. The original matched guise task, conducted by Lambert et al. (1960), examined bilingual French Canadians’ attitudes toward French and English and found that English was typically rated higher on dimensions such as intelligence, likability, and even height, even though it was the same speaker (and content) they heard in both the English and French guises. Since then, the matched guise technique has been developed to examine the social meanings attributed to individual variants or clusters of variants as well. Often this involves synthetic manipulation of the same exact sound file. One well-known example is Campbell-Kibler (2011) who found that a guise using sociolinguistic variant ING was rated more intelligent, more articulate, and less likely to be a student, while the IN guise was rated less formal and less likely to be gay. It has been found repeatedly in the sociolinguistic literature that linguistic variants signal social meaning.

Conversely, social information has been shown to affect linguistic processing as well. In a most basic example, unexpected social information can hinder language processing. Expectations about a speaker have been shown to be extremely important to speech perception, and when expectations do not align, individuals are in general worse at perceiving speech. A notable example is that, when a man utters a sentence with semantic content outside of what might be expected for a man (e.g., "I'm pregnant") or when a child utters a sentence with content outside of what might be expected for a child (e.g., "I just quit smoking"), event-related brain responses indicate surprisal as early as 200-300 ms (Van Berkum et al., 2008). Relatedly, McGowan (2015) finds that, when a Chinese-accented voice is played, participants are more accurate in transcribing the recording when presented with a Chinese face than with a Caucasian face.

Social information of many different types has been shown to alter linguistic perception in predictable ways beyond processing ease. For instance, beliefs about the region a speaker is from influence classification of linguistic variants. Niedzielski (1999) found that, when speakers believed they were listening to a Canadian speaker, they chose raised-diphthong tokens as representative of the /au/ diphthong, but when they thought they were listening to a Detroit speaker they did not, even though all speakers were listening to the same Detroit speaker; because the raised /au/ diphthong is stereotypically associated with Canadians but not with Detroit speakers (though both use it), participants perceived the token as more raised only when it fit their expectations based on speaker place of origin. Similar effects have been noted for beliefs about the age of a speaker. Similarly, D'Onofrio (2015) found that listeners were more likely to look at and click on TRAP (as opposed to LOT) category words after hearing an ambiguous word between TRAP and LOT, when presented with an image signalling "California" or "Valley-Girl," both of which are associated with TRAP-backing. Hay et al. (2006b) manipulated the age of a speaker via pictures presented during the experiment. When asked to judge vowels undergoing merger for younger speakers of NZE, participants were influenced by perceived speaker age, such that those who saw an image of a younger speaker were less able to distinguish between the vowels in a perception task. In a similar study, Koops et al. (2008) found that the perceived age of a speaker influences the degree to which participants believe a speaker participates in the PIN/PEN merger, which is receding among younger speakers. Speaker race has also been shown to influence linguistic perception. Staum Casasanto (2010) found that ambiguous sentences with potential t/d deletion were more often interpreted as a deleted variant when participants were told the speaker was African American. Gender is another feature known to influence perception. For instance, Strand (1999) found that a listener's belief about the gender of a speaker influences categorization of /s/ and /ʃ/.

Such effects have been shown to occur even when participants are not told to attribute a trait to the speaker—or even when no speaker is present at all. For instance, in a partial replication of Niedzielski (1999), Hay et al. (2006a) found that participants who were told that they were listening to a New Zealander were still influenced by the words "Australia" or "New Zealand" printed at the top of their answer sheet, in that they were more likely to hear a fronter /ɪ/ vowel in the Australia condition, reflecting the realization of this vowel in Australian English. Hay and Drager (2010) have suggested that even subtler nonconscious activation of regional groups can influence speech perception as well. For each experimental condition, a stuffed animal representative of either Australia (kangaroo and koala) or New Zealand (kiwi) was placed somewhere in the room, though the participant's attention was not drawn to it. This brief exposure to objects meant to simply *evoke* the concept of a particularly region was enough to influence participants' behavior on a vowel identification task. Though participants all heard the same vowels (which differ in Australian and New Zealand English) produced by a New Zealander, they classified vowels on a continuum differently depending on whether the concept of New Zealand or Australia was activated via these stuffed toys. Similarly, Sanchez et al. (2015) found that speakers in a corpus produced more Australian-like variants of KIT and TRAP (but not DRESS) when talking about Australian topics, and experimental participants showed the same effect after producing Australia-related lexical items. Relatedly, Love and Walker (2013) found that soccer fans in the UK became more r-ful when talking about American football. These studies all show that social information (even quite abstract social concepts) is able to influence both linguistic perception and production.

While this recent work has demonstrated that linguistic and social knowledge bidirectionally influence one another, little is known about the way social factors are mentally represented and how social representations are linked with linguistic ones. For instance, What types of social categories are mentally represented and available to be linked to linguistic categories? How are these links activated and used in communication? How do perception and production reflect these links in similar or different ways? One of the key goals of the dissertation is to shed light on the mental relationship between social and linguistic knowledge, as well as the relationship between linguistic perception and production, using linguistic convergence as a test case.

1.2 Convergence

Linguistic *convergence* is when language users shift their speech to become more similar to another speaker. This is part of the broader process of linguistic *accommodation,* which encompasses *convergence* behaviors, as well as *divergence* (when people shift their speech *away* from an interlocutor), and maintenance (when people maintain their own speech norms despite conversing with an interlocutor with different speech patterns). This process has many different names within linguistics, social psychology, and cognitive science, including *alignment, spontaneous imitation,* and *entrainment.*

The dissertation focuses on linguistic convergence as a window into the relationship between linguistic and social knowledge for several reasons. This phenomenon is an ideal test-case for probing this relationship because convergence has been reported to occur in response to both linguistic and social cues, and it comprises both (arguably more automatic) perceptual behaviors and (perhaps more deliberate) production behaviors—-allowing us to observe the influence of social information in both production and perception. Further, convergence has broader implications for many important questions in linguistics, including the relationship between perception and production and the mechanisms behind language change.

In order to better control for both observed and expected linguistic behavior, the dissertation investigates convergence using a series of experiments, which has shown to be similar in a number of ways to convergence in more naturalistic settings. Convergence has been observed in both naturalistic conversation (Giles et al., 1991; Bourhis and Giles, 1977; Natale, 1975a; Pardo, 2006), and laboratory experiments such as speech shadowing tasks (Babel, 2012,0; Goldinger, 1998; Goldinger and Azuma, 2004), in which participants read target words, then repeat the same target words after a model talker, and pre- vs. post-exposure productions are compared. While these types of data are quite different in a number of respects, both have yielded comparable results. For instance, in a comparison of convergence in naturalistic conversation and a laboratory speech shadowing task, Pardo et al. (2018) found comparable levels of convergence in both and a weak correlation in degree of convergence between the two contexts.

It is necessary here before delving into the literature on linguistic convergence to first establish some relevant terminology and theoretical distinctions important for understanding the phenomenon of linguistic convergence. First it is necessary to draw a distinction between the *trigger* and the *target*, both of which are necessary components of convergence. We can think of the *trigger* as anything that cues an individual to converge in the first place, while the *target* is what the individual intends to produce as a means of converging. These are often assumed to be the same; usually, when convergence is reported, it involves hearing a particular feature (the trigger) that is then imitated (the target). However, I draw a crucial distinction here that this need not be the case, and argue the opposite point that the trigger and the target of convergence can be quite distinct.

I propose a terminological and theoretical distinction differentiating two types of convergence that vary primarily in the relationship between the trigger and target. First, is the more traditionally cited type of convergence, which requires a linguistic target that is triggered by the same linguistic form[1] observed in real time. An example would be producing lengthened VOT for voiceless stops after hearing a talker produce voiceless stops with lengthened VOT. I refer to this as ***input-driven*** convergence because the target is directly derived from the input. I contrast this with what I refer to here as ***expectation-driven*** convergence, which entails converging toward a previously heard and later recalled target that is not locally observed. The trigger of convergence in these cases can be linguistic, but need not be, and the target is derived from expectations based on pre-existing knowledge. An example of expectation-driven convergence would be producing a linguistic form associated with Southern speech (for example, *y'all*) when conversing with a Southerner who never actually uses this form. The idea is that *y'all* is stereotypically associated with Southern speech, and an individual may converge toward this expected behavior in the absence of any real evidence that a particular speaker uses this form. Though such *expectation-driven* convergence is theoretically possible and has been cited anecdotally (e.g., Bell, 2001), the phenomenon has not been established in a controlled experimental study. One of the primary goals of this dissertation, however, is to design a suitable paradigm for eliciting expectation-driven convergence to determine (1) whether it occurs and (2) the nature of its occurrence.

The following subsections provide an overview of the current state of knowledge on linguistic convergence, focusing especially on areas that have informed the research questions, experimental

[1] *Linguistic form* can refer to linguistic units at basically any level of linguistic structure, including but not limited to the acoustic properties of phonemes, syntactic structures, or lexical usage, as my definition of input-driven convergence makes no assumptions about the level of linguistic structure at which convergence operates or the extent to which convergence generalizes across lexical items or phonemes, for instance.

design, and predictions of the dissertation. Section 1.2.1 reviews findings on the target of convergence, focusing on the extent to which the target of convergence has been shown to abstract away from the trigger. Section 1.2.2 outlines proposals for the mechanisms responsible for linguistic convergence and how these relate to the two types of convergence—expectation-driven and input-driven—proposed above. Section 1.2.3 then reports previous findings on various factors that have been shown to facilitate (or hinder) linguistic convergence, including speaker attitudes, feature salience, and individual cognitive traits.

1.2.1 The target of convergence

The crucial difference between input-driven convergence and expectation-driven convergence is the extent to which the *target* of convergence can generalize away from the *trigger*. In the simplest conceptualization of convergence, the trigger and the target are the same. That is, an individual may hear their interlocutor produce *pancake* with lengthened VOT on /p/, then go on to produce the word *pancake* with lengthened VOT on /p/. However, we can imagine that the target of convergence might generalize to some extent, such that other words besides *pancake* are produced with lengthened VOT. For instance—this individual may converge by producing *poodle* with lengthened VOT for /p/, or maybe even generalize further and produce *kitten* with lengthened VOT for /k/. We may, then, wish to think of input-driven and expectation-driven convergence not as categorically different phenomena, but as two points on opposite ends of a gradient scale. Convergence to the same lexical item previously heard from an interlocutor would be maximally input-driven—but what about convergence to the same phoneme but in a novel lexical item? What about to a different but related phoneme? The trigger may even be so far removed from the linguistic target that it is not linguistic at all. There are in theory varying degrees to which convergence could generalize, and the question of the extent to which convergence generalizes is not new in convergence research. The literature on this topic is be reviewed here.

We know some things about the degree to which the target of convergence can generalize away from the trigger, though there have been some conflicting findings. For one, there have been several claims in the literature that suggest that the target of convergence is quite narrow and unable to generalize. For instance, Goldinger (1998); Goldinger and Azuma (2004) argue that phonetic convergence does not extend beyond the lexical level, based on findings that speakers imitated various properties of lexical items they have heard, but imitation did not generalize to novel lexical items. Other studies have contradicted this claim with evidence that imitation is in fact generalized beyond the word level. Pardo (2006) for instance, found that speakers' vowel productions became more similar to their conversational partners'. Crucially, these productions consisted of words that participants did not hear in the exposure phase, suggesting that convergence can be generalized at the phonemic level across words.

Others have provided evidence that phonetic convergence can even generalize across phonemes. Importantly, Nielsen (2011) found that imitation of sub-phonemic features can be generalized to new lexical items and even new phonemes. Specifically, she found that, when participants were exposed to artificially lengthened VOT for /p/-initial words, not only did participants imitate artificially lengthened VOT for new /p/-initial words, but they also produced lengthened VOT for new /k/-initial words, though the effect was somewhat weaker. She proposed that the target of imitation in this case may be at the level of the shared [+ spread glottis] feature, though the mechanism that allows for such generalization is not apparent. Recent work by Wilson et al. (2016), for instance, questions this assumption that the target of convergence might be featural. They suggest that, perhaps imitation in the Nielsen (2011) study did not occur at the featural level, but rather at the level of the speaker. They found that mean VOT of voiceless stops covaries within speakers. This means that the speakers may have actually produced lengthened VOT of /k/-initial words when they heard lengthened VOT of /p/-initial words because they were accessing knowledge that VOT of /k/ and /p/ generally covary within speakers. This plausible explanation of Nielsen's findings is supported by other empirical results suggesting that listeners know that longer VOT of /k/ is more likely to be produced by a speaker who also produces longer VOT for /p/, even if the listener has never heard the speaker produce a /k/. For instance, Theodore and Miller (2010) found that listeners chose longer variants' of /k/ as representative of a talker who recently produced lengthened VOT for /p/, and shorter variants of /k/ for a talker who recently produced shorter VOT for /p/.

Zellou et al. (2017) have also recently suggested that listeners are influenced by a more holistic model of the talker in shadowing tasks, as opposed to individual instances of the linguistic form itself. They found that participants who heard a hyper-nasalized speaker in the first block of shadowing increased their degree of coarticulatory nasalization. However, if hyper-nasalization was heard in a second block (that is, after an initial block of regular nasalization) participants' degrees of nasalization leveled out as if they were averaging the nasalization across all tokens they had heard from the speaker and converging toward that average. They ultimately suggest that participants may imitate isolated phonetic forms immediately after hearing them, but after a decent amount of exposure (or a delay between exposure and their own productions), they converge toward a holistic model of the speaker based on accumulated utterances, rather than only toward the most recent tokens. Such findings have provided evidence that convergence can generalize to a greater extent than previously thought.

The idea that speakers may converge more holistically, rather than toward specific individual targets, is not new. In fact, various literature rooted in Communication Accommodation Theory (CAT) has proposed that speakers often shift their speech to align with some abstract idea of whomever they are speaking with. For instance, Auer and Hinskens (2005) advocate for an "Identity-Projection Model" of convergence, which suggests that instead of converging toward "observable behaviour of the recipient ... speakers converge to a stereotype of the 'model' receiver, not the actual partner in direct communication" (p. 341). That is, speakers may change the way they speak to match how they believe their conversational partner speaks, regardless of whether their conversational partner actually speaks that way. As evidence, they cite Bell (2001), in which an Anglo interviewer uses the tag *eh* frequently when conversing with a Maori interviewee but not when conversing with an Anglo interviewee. Although the Maori speaker did not use the tag *eh,* it is a stereotypical feature associated with Maori speech. This is given as (admittedly somewhat anecdotal) evidence that speakers shift their speech to align with pre-existing beliefs about what an interlocutor should sound like. This idea of convergence triggered by the speaker rather than the speech itself can be traced further back to Thakerar et al. (1982), who terms accommodation which "responds to what the speaker mistakenly assumes will be the addressee's speech on the basis of the addressee's nonspeech attributes" to be "subjective accommodation" (Bell, 1984, p. 168). Earlier models of convergence in sociolinguistics such as Communication Accommodation Theory and Audience Design suggest that ideas about an interlocutor's identity are equally—if not more—important in eliciting convergence than the actual linguistic forms themselves. While concepts like the Identity Project Model and Subjective Accommodation are intuitively appealing, they currently lack a firm empirical foundation, which is why a central goal of the dissertation is to test such phenomena using a series of more controlled experiments.

1.2.2 Mechanisms of Convergence

There have been several models proposed to account for linguistic convergence, which can be separated into two broad categories: (i) automatic accounts of convergence which rely on tight perception-production links where production targets are derived directly from perception, and (ii) more controlled or intentional accounts that maintain that convergence is motivated primarily by external (usually social) factors. Broadly speaking, automatic models of convergence only account for input-driven convergence, while accounts focusing on the socio-psychological motivations for convergence may allow for a wider range of convergence triggers, including non-linguistic information such as social categories (i.e., expectation-driven convergence). This section reviews several of the better known proposed mechanisms of linguistic convergence.

There are a few different proposed convergence mechanisms that could be classified as automatic. The first relies on exemplar-based representations, which are derived from linguistic perception, and from which production targets are drawn (e.g., Goldinger, 1998; Goldinger and Azuma, 2004). Under such a model, episodic lexical items are stored in memory with detailed phonetic information attached to each. "Echoes" or memory traces of previously heard lexical items are activated during production of the same lexical item, and production targets are derived from the distribution of these activated traces. Convergence occurs when the distribution from which production targets are drawn has shifted to encompass recently heard forms. There are a number of distinct pieces of evidence that support such a model. For instance, in Goldinger (1998); Goldinger and Azuma (2004), imitation only occurred for identical lexical items (though note that later studies have not replicated lexical specificity in convergence), providing support for an exemplar model which is lexically specific. Secondly, imitation was found to be greater for lower frequency words (Goldinger, 1998), which supports an episodic usage-based model in that the fewer traces there are to compete with the prime words that listeners heard in the experiment, the more influential the memory traces of these words would be. Another proposal for an automatic model of linguistic accommodation is the Interactive Alignment Model (Pickering and Garrod, 2004), which suggests that individuals align their linguistic representations due to automatic priming mechanisms that serve to increase intelligibility for both speakers. Under this model, linguistic convergence is viewed as the result of automatic structural priming mechanisms that align on all levels of linguistic structure.

Relatedly, Pickering and Garrod (2013), Fowler et al. (2003), and Sancier and Fowler (1997) appeal to a general tendency to imitate gestures, stemming from automatic priming of perceived motor gestures, as a mechanism of convergence, an idea rooted in Motor Theory (Liberman and Mattingly, 1985). Under one account, "actors construct forward models of their actions before they execute those actions, and that perceivers of others' actions covertly imitate those actions, then construct forward models of those actions" (Pickering and Garrod, 2013, p. 36). However, most automatic models have been criticized for failure to allow for external factors to influence production (e.g., Brennan and Metzing, 2004; Brown-Schmidt and Tanenhaus, 2004; Krauss and Pardo, 2004). As Pardo et al. (2012) notes, "it is unclear how an automatic priming mechanism can incorporate the sorts of situational modulations that have been observed in conversational phonetic convergence. That is, if a perceiver automatically resolves the articulatory gestures that a talker made while producing an utterance, then production should follow suit and result in maximally converged phonetic forms" and concludes that, "because talkers do not match more precisely or consistently and their departures from matching are not due to random variability in perception and production, speech perception and production during conversational interaction are not well-explained by automatic priming mechanisms" (p. 758). Divergence effects have also been difficult to account for without appealing to (at least partially) socially-motivated models of convergence.

Those who argue for more intentional motivations for convergence tend to appeal to the influence of social factors. For one, some groups have been found to converge more than others. For instance, the role participants were assigned to in a map task has been shown to influence degree of accommodation (Pardo, 2006; Pardo et al., 2010). If automatic mechanisms were primarily responsible for convergence, it is unclear why different groups would exhibit differing degrees of convergence. Further, the fact that perceived social traits of and attitudes toward the interlocutor modulate convergence effects suggests social motivations. Babel (2009) found that vowel imitation was affected by implicit social measures of how the participant felt abut the talker including ratings of likability and attractiveness. Babel (2010) also found that participants' degree of convergence correlated with their social biases toward entire dialect groups, with speakers who scored highly for pro-Australia bias converging more to an Australian speaker. Similarly, Balcetis and Dale (2005) found that syntactic alignment occurs to a confederate who acts nice (compared with a mean confederate) and annoyed (compared with a patient confederate), perhaps as a means of communicative repair. These findings may seem counter to one another, which highlights the fact that we don't as of yet have a good idea of what to expect in regard to how social factors precisely influence accommodation, though there is a good deal of evidence that they play *some* role.

The role of social factors is particularly apparent in the sociolinguistic literature that treat accommodation as a form of style shifting that occurs in response to others present in the conversation. Communication Accommodation Theory (CAT), for instance, assumes that the primary motivation for shifts in speech is to win approval and manage social distance (Giles et al., 1991). CAT also accounts for non-convergence behaviors such as speech divergence and maintenance, as well as for hyper-convergence, where a speaker overshoots the convergence target (Giles, 1980). As Bell (1984) notes, accommodative style shifting is necessarily social, as "Variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the 'social' dimension" (p. 151). Bell's Audience Design Model in a sense extends CAT to allow for speech to be influenced by not only one's primary addressee, but also by others who may be involved in the communicative situation (auditors, overhearers, eavesdroppers), as well as other nonpersonal variables such as topic and setting. His model of audience design assumes that, when people speak, they take into account primarily those who would be listening to that speech. Note also that a few recent studies have found shifts in participants' speech in response to social concepts, without an explicit interlocutor (e.g., Drager et al., 2010; Sanchez et al., 2015; Love and Walker, 2013). Additionally, addressee and topic-based style shifting have also been observed in the literature (Bell, 1984; Love and Walker, 2013), though style-shifting is generally considered to be shifting along a single spectrum of formality within a single dialect. It is unclear as of yet how these phenomena relate to knowledge-driven convergence.

The influence of social factors on linguistic convergence has often been taken as evidence that speakers have (at least some) control over their accommodative behavior. There is a commonly held view that social motivation requires awareness, and that behaviors we are not aware—or in control—of can therefore not be social. Labov (1963), for instance, concluded about the centralization of /aɪ/ and /au/ in Martha's Vineyard that "centralized diphthongs are not salient in the consciousness of Vineyard speakers. They can hardly therefore be the direct objects of social affect." (p. 40). However, recent research has questioned this assumption, arguing that it is not a given that social motivations for linguistic behavior must mean that speakers are consciously controlling or even aware of that behavior. Campbell-Kibler (2016), for instance, argues that "social" does not necessitate consciousness, providing evidence that many social cognition behaviors observed must occur very rapidly and without conscious awareness. Particularly in light of the findings that social information influences linguistic perception, sociolinguistic associations have been considered to occur more rapidly and automatically than previously thought. As Campbell-Kibler (2016) notes, "sociolinguistic cognition is a kind of cognition" and "many important processes, including social processes, at least occasionally occur quickly, without introspective awareness and/or in ways apparently at odds with verbally reported or experimentally instructed intentions" (p. 10-11). In fact, recent studies on social cognition have found an influence of social information in linguistic processing is as little as 200-300 ms.

It remains unclear at this point whether a single mechanism will ultimately be able to account for all facets of linguistic convergence. For instance, we know that speakers converge toward variables that attract less social awareness (e.g., nasalization in Zellou et al. (2017)), but could expectation-driven convergence also occur to these variants? What types of non-linguistic information would even be able to trigger convergence in the dimension of hyper-nasalization? Whether input-driven and expectation-driven convergence are actually governed by the same mechanisms is an empirical question that has yet to be answered. Rather than asking whether automatic accounts or those which allow for the integration of social information are better suited to account for either type of linguistic convergence, I would argue that a better question to ask is *how* social information modulates these perception-production links in accommodative behavior.

1.2.3 Factors Facilitating Linguistic Convergence

Studies on convergence have shown a great deal of individual variation. Some individuals may converge toward a linguistic variant spoken by a given talker in a particular setting, while under the same parameters others will converge less, or not at all. That is, even given an identical experimental design, not all individuals behave in the same way (e.g., Babel, 2012; Pardo, 2006; Nielsen, 2011). This section reviews the literature on various factors that have been shown to facilitate convergence, which may begin to explain the differences in how individuals respond to the same experimental stimuli. Specifically, this section investigates the role of various factors in convergence behavior, including demographic variables such as gender, motivation to converge including attitudes toward the model talker or language variety, experience with the language variety, and individual differences in cognitive processing style and personality traits, measuring domain-general capacities, which have been proposed as potentially indicative of linguistic behaviors such as phonetic flexibility (e.g., Yu et al., 2013).

Experience vs. Novelty

Perhaps most obviously, experience with a given language variety would be expected to influence whether an individual converges to that variety. In fact, studies have found that shifts in speech toward a given language variety are in fact dependent on linguistic experience. For instance, Love and Walker (2013) found that exposure to both American and British English contributed to shifts toward greater rhoticity when discussing American football vs. English soccer. Similarly, Sanchez et al. (2015) found that New Zealanders shifted to produce more Australian-like variants, and that greater experience with Australian English led to larger shifts, but only for the less salient DRESS vowel, suggesting that experience may matter most for variants that are not stereotypically marked or salient.

Indeed, salience has been shown to play a role in convergence. Babel et al. (2014), for one, found that speakers converged more toward voices judged as atypical than for typical sounding voices, suggesting that the novelty of the voice led to more attention paid to that talker. The same holds for the salience of individual linguistic targets. Drager et al. (2010) found that New Zealand English speakers shifted their speech to produce more Australian-like variants after simply being exposed to facts about Australia, and the effect was greater for the more socially salient KIT vowel than for the DRESS vowel. Not all studies have supported these claims that more socially salient variables show greater convergence effects. For instance, Babel (2010) and Walker and Campbell-Kibler (2015) both found that speakers converged toward Australian speakers' DRESS vowel, which is thought to be below the level of conscious awareness, but found no convergence toward the Australian KIT vowel, of which speakers seem to have some sociolinguistic knowledge, which directly contrasts with the findings of Drager et al. (2010), who found a greater degree of shift for the KIT vowel than for the DRESS vowel. However, in the Drager et al. (2010) study, similar to the expectation-driven approach of the dissertation experiments, the shift in production was expectation-triggered, and participants didn't actually observe Australian vowels before converging toward them. This could suggest that salience would be more important for facilitating convergence when an actual linguistic target is not locally present, as it may be more strongly encoded and remembered in the absence of a phonetic target.

The role both experience and salience play in convergence may appear to contradict each other. The more perceptually salient something is, the more likely it is to be noticed, and therefore the more likely convergence is. At the same time, experience with a variant would contribute toward greater capacity for producing that variant (i.e., individuals cannot target a variant they have never heard). Indeed, contradictory results have been reported for the role of dialect similarity in convergence. Some have shown that greater similarity between dialects leads to greater convergence. For instance, Kim et al. (2011) found that speakers converged more to interlocutors with the same dialect background. Similarly, Babel (2009) found that California speakers converged most to low vowels and argues that this is because the low vowels show the most intra-speaker variability, and the model speaker's targets were therefore within the participants' production repertoires. However, not all studies have produced congruent results. Other studies have shown that speakers converge more to novel variants outside of their own dialect. For instance, Walker and Campbell-Kibler (2015) actually observed *more* convergence from speakers to dialects most different than their own. Midland speakers converged more to New Zealand speakers than to Inland North speakers, and New Zealand speakers converged more to U.S. dialect than to Australian English. Additionally, the vowel that showed the most convergence across dialects—the DRESS vowel—is the vowel that is the most phonetically different across dialects, although it is considered to be below the level of conscious awareness. These findings, in conjunction with seemingly contradictory findings about a tendency to converge within one's repertoire, have led Walker and Campbell-Kibler to conjecture that speakers shift toward the biggest and most salient differences, but only if their repertoires allow it.

Affective measures

Attitudes toward specific talkers or language varieties have also been shown to influence degree of convergence. The extent to which a speaker "likes" an interlocutor or model talker is perhaps the most frequently discussed affective facilitator of linguistic convergence. Intuitively, this is not surprising. Indeed, Dijksterhuis and Bargh (2001) argue that the simplest factor influencing whether perception translates into behavior is liking. While "likability" can mean many things, this is often measured in terms of perceived attractiveness, following from the similarity attraction hypothesis (Byrne, 1971), which claims that individuals try to be more similar to those they are attracted to.

This intuition has been somewhat borne out in the convergence literature. Babel (2012), for instance, found that vowel quality imitation was subtly influenced by how attractive participants judged the model talker. Similarly Babel et al. (2014) found that participants converged to a greater extent when listening to voices previously judged as more attractive. In both cases, this finding held only for female participants. It has also been suggested that positive feelings generally facilitate convergence. Babel (2010) also found that participants' degree of convergence correlated with their social biases toward entire dialect groups, with speakers who scored highly for pro-Australia bias converging more to an Australian speaker's vowels. Similarly, Drager et al. (2010) found that New Zealanders shifted their production of the KIT vowel to a more Australian-like variant, and the shift was greater when they were previously presented with good facts (as opposed to bad facts) about Australia. However, the pattern was reversed for sports fans, who shifted less in the "good-facts" condition, relative to the "bad-facts" condition, which they explain as a defensive attitude regarding being a New Zealander after hearing good facts about Australia, since NZ sports fans typically have negative attitudes toward to begin with. Even more generally, Yu et al. (2013) found that participants converged toward VOT more when they heard a story from the model talker with a positive (rather than negative) outcome. Attitudes also seem to play a role in divergence. In a seminal study, Bourhis and Giles (1977) found that Welsh speakers who were more invested in Welsh language and culture diverged from their interlocutor and become more Welsh accented during an interview with a non-Welsh speaker who expressed negative opinions on the vitality of the language.

Findings regarding the role of attitude in convergence, however, are not always so straightforward. As mentioned above, Balcetis and Dale (2005) found that syntactic alignment occurred when the confederate acted nice (compared with a mean confederate), but also when they acted more annoyed (compared with a patient confederate). The authors proposed that convergence may be used as a means of communicative repair, which would explain the increased rates of convergence toward a confederate who acted annoyed. It may also be the case that, rather then being driven by straightforwardly positive or negative affect, convergence is stems from more nuanced goals that may differ depending on the circumstances.

Cognitive Processing Style and Personality Traits

One individual differences dimension that has received some attention recently for its potential explanatory role in understanding individual differences in linguistic behavior is cognitive processing style, and personality traits by extension. Cognitive processing styles refer to an individual's habits in processing, encoding, and retrieving information, which contribute to the development of stable personality characteristics (Witkin et al., 1954).

Domain-general differences in individual cognitive processing styles and personality dimensions have been hypothesized to contribute to individual differences in linguistic behavior. Recent studies have been successful in linking individual differences in aspects of phonetic production and perception to domain-general cognitive factors, including in tone learning (Perrachione et al., 2011; Hedger et al., 2015), VOT production (Lev-Ari and Peperkamp, 2014), perceptual compensation (Yu, 2010; Dimov et al., 2012; Kingston et al., 2015), the Ganong effect (Stewart and Ota, 2008), phonetic imitation (Yu et al., 2013; Lewandowski and Jilka, 2019), fast speech reduction (Turnbull, 2015), and auditory word recognition (Arciuli and Cupples, 2003; Janse and Newman, 2013; Lev-Ari and Peperkamp, 2014; Baese-Berk et al., 2015).

One domain-general dimension in which individuals might differ that may contribute to convergence is in their desire to be perceived in a socially desirable way. This is potentially an important factor for convergence, particularly under a Communication Accommodation Theory model (Giles et al., 1991), in which convergence is a tool used for managing social distance and signaling social affiliation. In fact, Natale (1975a) found that scores on the Marlowe-Crowne Social Desirability Scale—which measures the degree to which individuals respond to self-assessment surveys in perceived socially desirable ways—predicted convergence to vocal intensity in dyadic communication (Natale, 1975a), as well as pause switching duration (Natale, 1975b).

Other traits related to a desire for social approval have been reported to correlate with convergence. Aguilar et al. (2016), for instance, found that individuals with a high level of trait rejection sensitivity (the tendency to expect social rejection) exhibited greater convergence during conversation than those with low trait rejection sensitivity. Lewandowski and Jilka (2019) similarly found that Neuroticism scores on the Big Five Inventory (BFI) of personality traits and lower Behavior Inhibition Scale (BIS) scores (which measures motivation to avoid aversive outcomes) correlated with convergence in a cooperative Diapix task, in which partners try to find the differences between the two pictures each one has by communicating with each other. They interpret greater neuroticism as signalling a higher need for social approval, which would facilitate convergence under a CAT model. Further, they suggest that those with lower behavior inhibition scores are more likely to try converging because they have a lower punishment sensitivity.

Various measures of attention, which may affect the level of engagement with exposure materials and greater attention paid to fine phonetic details, have also been cited as important facilitators of linguistic convergence. Yu et al. (2013), for instance, found that individuals with greater Openness and Attention Switching scores on the Big Five Inventory and Autism Quotient surveys, respectively, imitated lengthened VOT to a greater extent. They suggest that Openness may be a proxy for greater perceptiveness and engagement with the narrator's speech, and Attention Switching also indicates greater focus, both of which potentially increase attention paid to exposure materials, which has been shown to facilitate convergence. Lewandowski and Jilka (2019) also found greater convergence for those with higher Openness scores, as well as for those with greater Attention Switching, though this was measured more objectively using the Simon Test, rather than the self-assessed AQ survey.

Taken together, heightened focus/attention and a greater desire for social approval appear to be the primary cognitive/personality traits proposed as important facilitators of linguistic convergence. While a greater desire for social approval would similarly be predicted to facilitate expectation-driven convergence (perhaps more so than for input-driven, because it is less straightforwardly understood as a potentially automatic consequence of perception), it is unclear whether measures of attention and focus would similarly contribute to convergence when there is no local phonetic target for participants to focus their attention on. Based on these findings, we might predict that expectation-driven convergence would be more influenced by social approval measures than by attention measures.

Demographic Factors

The only demographic factor that has been consistently investigated as a facilitator of convergence is speaker gender, perhaps due to the belief that females exhibit increased affiliative strategies (Giles et al., 1991). However, findings have been largely contradictory. One of the first studies reporting such an effect is Namy et al. (2002), who found that females converged more than males in a shadowing task. However, this was only the case for one out of four model talkers, and both males and females converged equally to the other three model talkers. Babel et al. (2014) also reported greater convergence for females in a shadowing task, though this again varied across model talkers. Miller et al. (2010) also found the females converged more to two model talkers in same-sex pairs, but the patterns was not replicated in a later experiment. Pardo (2006) and Pardo et al. (2010) found that males converged more than females in conversational interaction, though others have found no effect of speaker gender (Pardo et al., 2013; Yu et al., 2013; Lewandowski and Jilka, 2019). More recently, Pardo et al. (2017) found that females did converge more, but only with lower frequency words, and suggests that previous studies finding higher rates of convergence among females crucially used only lower frequency items. The role of demographic factors such as speaker gender are therefore still largely unclear.

Chapter 2

Project Design

The dissertation employs a series of production and perception experiments in order to accomplish two main goals. The first goal is to establish expectation-driven convergence as a real and replicable phenomenon, observable in a controlled experimental study. In line with this goal, the dissertation develops an experimental paradigm suitable for eliciting expectation-driven convergence. The second goal of the dissertation is to better understand what drives this behavior, including its perceptual underpinnings, triggers, and relationship to input-driven convergence. Ultimately, accomplishing these two main goals will lead to a better understanding of the relationship between social and linguistic knowledge. This chapter lays out the design of three experiments designed with these goals in mind. Here, I present a general descriptions of each experiment's goals, state the research questions and predictions guiding each experiment, and explain how the three experiments fit together to examine the relationship between linguistic and social knowledge in both speech perception and production, using convergence to glide-weakened /aɪ/, a salient feature of Southern U.S. English as a test case.

The general experimental paradigm is the same for all three experiments, which elicits expectation-driven convergence toward Southern speech using a novel Word Naming Game task. The task is designed to elicit specific tokens of /aɪ/ from participants, before, during, and after exposure to a model talker, without participants ever hearing this vowel from the model talker. The task utilizes a between-subjects design where participants are randomly assigned to either the Southern or Midland (control) model-talker conditions. In this task, participants are presented with "clues" describing target items, and are tasked with guessing the word described by the clue by naming it aloud in the phrase "The word is [Blank]." In the Baseline Phase (Phase 1), participants read the clues to themselves before giving their responses aloud, so that baseline speech production can be measured prior to exposure from the model talker. In the Exposure Phase (Phase 2), clues are presented auditorily by either a Southern or Midland (control) model talker. In the Post-Exposure Phase (Phase 3), participants go back to reading clues from the screen so that the longevity of any shifts in speech production can be assessed. I then acoustically compare the elicited /aɪ/ vowels across these phases to determine whether convergence has occurred. Crucially, in order to assess whether convergence occurs toward expectations alone, the model talker never uses the /aɪ/ vowel throughout the course of the experiment. In every experiment, half of the participants grew up in the South while the other half grew up outside of the South so that in order to investigate the role of dialect experience in convergence. In order to collect data from a large number of participants (N=478) across a wider demographic range, the experiment was administered online and audio data was collected through participants' computer microphones. This general design is modified across the three experiments to target specific questions, outlined below, and the distribution of participants across the experiments is shown in Table 2.1.

2.1 Experiment 1

Since the three experiments all use the same general design to elicit the same phenomenon, it is important to first establish that the experimental paradigm works and the phenomenon we are testing is real, replicable, and to some extent generalizable. To accomplish this, I first ran an initial pilot version of Experiment 1 intended to (1) test the paradigm for its ability to elicit convergence, (2) pilot clues and omit or modify any that resulted in high error rates, and (3) determine which acoustic measure of /aɪ/-glide weakening is the most accurate and precise measure of production shifts triggered by the experimental manipulation, isolated as best as possible from global speech shifts across the course of the experiment due to factors such as fatigue. Experiment 1 is the most basic version of the Word Naming Game task (see Figure 2.1), with the goal of simply determining (1) whether expectation-driven convergence occurs and (2) whether the task is sufficient in eliciting expectation-driven convergence. Specifically, the research question for Experiment 1 is: Do participants converge toward a Southern talker by producing more glide-weakened /aɪ/ tokens, a salient feature of Southern speech, after exposure to a Southern talker who produces no tokens of /aɪ/ throughout the course of the experiment? Based on anecdotal evidence that expectation-driven convergence occurs in naturalistic interaction (e.g., Bell, 1984), experimental evidence that expectation-driven convergence occurs in an artificial language toward alien species who serve as different social groups (Wade and Roberts, 2020), and the social salience of /aɪ/ glide-weakening (Labov, 2010), I predicted that I would indeed find experimental evidence of expectation-driven convergence toward Southern glide-weakened /aɪ/.

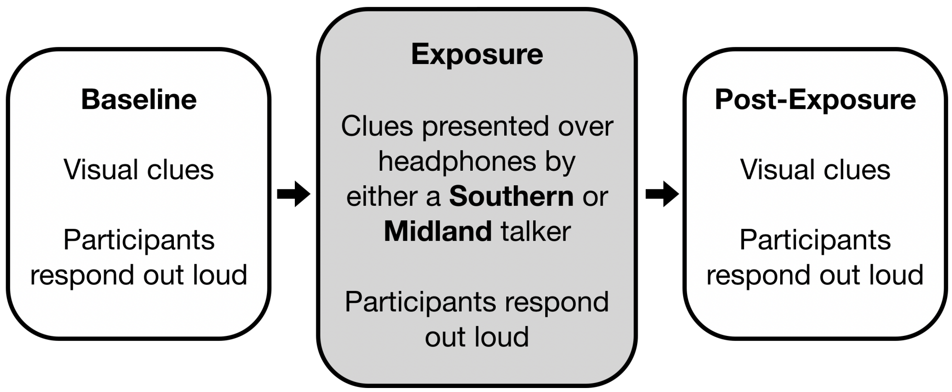

Figure 2.1: Experiment 1 Design

To test this hypothesis that participants will produce more glide-weakened /aɪ/ after exposure to a Southern-shifted talker who never produces this vowel, the design of Experiment 1 is as controlled as possible. Participants are randomly assigned to one of the two experimental conditions, and the phase in which tokens are elicited is counterbalanced across participants. The order of elicited items is also randomized across participants. This initial experiment, further, contains no perceptual components (as in later experiments), which may influence production data in unexpected ways. Additionally, a Midland voice (control) Condition, as well as a post-exposure phase are included to ensure any shifts we see from the baseline to the exposure phase are truly the result of the experimental manipulation. The full-scale version of Experiment 1, with some minor design alterations informed by the pilot results, supports this hypothesis, replicating the results of the pilot with a new subject pool, different model talkers, and an online (opposed to in-lab) presentation; this suggests that the experimental method is successful in eliciting expectation-driven convergence under somewhat different circumstances, and that the phenomenon of expectation-driven convergence is real and replicable. Beyond testing the hypothesis that expectation-driven convergence occurs toward Southern-speech, other investigations within Experiment 1 are largely exploratory and used in informing the experimental design and predictions of subsequent experiments. For one, Experiment 1 examined a range of possible individual differences and attitudinal measures that may be relevant in predicting convergence, which are then examined in detail in Experiment 2. The results of Experiment 1 are reported in Chapter 3.

2.2 Experiment 2

Experiment 2 uses an individual differences design, in which each participant completed separate perception and production tasks, along with a battery of surveys gauging attitudes, cognitive-style, and personality traits, which allows for perception and production to be directly compared on an individual—rather than group—level. Individual differences measures are then examined as individual predictors of perception and production behavior. There are several reasons for using an individual-differences design in Experiment 2. For one, examining perception-production links at the individual level is important because any relationship between perception and production would exist in the minds of individual speaker-listeners, and group-level examinations of this relationship may reflect the influence of shared community input rather than a true perception-production link. In other words, as Schertz and Clare (2019) note about the perception-production relationship: "group-level correspondences between the modalities cannot be taken as evidence for a causal link" (p. 9). Regarding the relationship between perception and production, I predict to find some evidence of an individual link. Precise alignment of perception and production would support a more automatic mechanism of convergence in which production shifts are an automatic consequence of perception. However, I predict to find a messier relationship between the two modalities, which may indicate only a partially automatic account that allows for the inhibition or enhancement of automatic behavior via social and attitudinal influences, and reflects the influence of external factors such as task-effects and fatigue, which may introduce variability (particularly in production) that could obscure this relationship to some extent. The second advantage of an individual-differences approach is that it allows for probing the individual-level predictors of input-driven and expectation-driven convergence, which may shed light on whether input-driven convergence and expectation-driven convergence are governed by similar mechanisms. Based on previous findings that many of the individual difference measures found to correlate with convergence have to do with *noticing* phonetic detail (which would presumably be relevant for input-driven convergence but less so for expectation-driven convergence), I predict that there will be some differences in individual predictors of these two types of convergence. Taken together, comparison of the three main components of Experiment 2 (perception, production, and individual differences measures) at the individual level will shed light on (1) the relationship between perception and production generally, (2) the perceptual underpinnings of convergence, and (3) the relationship between expectation-driven and input-driven convergence.

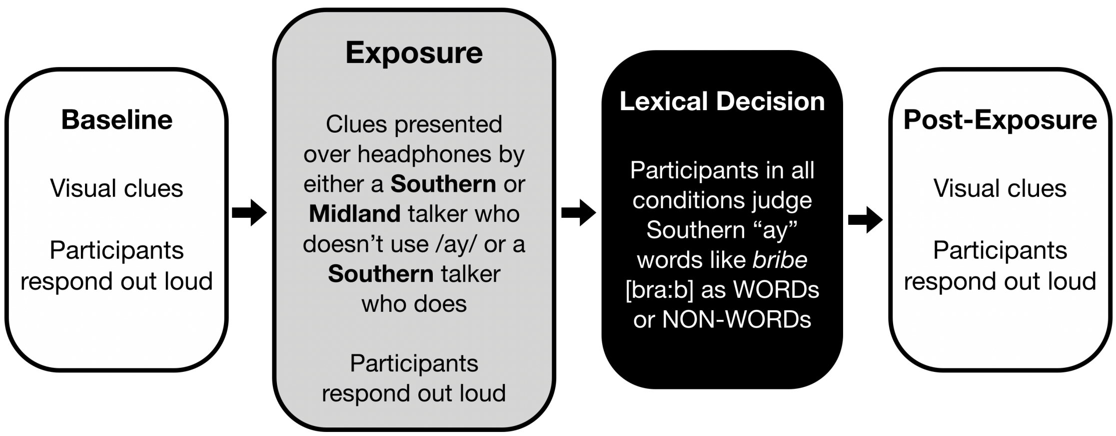

Figure 2.2: Experiment 2 Design

In an individual differences design, rather than counterbalancing item order with the goal of determining whether or not a general effect of the manipulation is present (i.e., if the null hypothesis can be rejected), elicitation of items is held constant across participants so that individuals can be more reliably compared to one another. Since all items are elicited in the same order for participants, it is possible that item-effects may amplify or diminish the main convergence effect. While this doesn't allow for probing whether an effect in general occurs, it does allow for comparison of *relative* shifts across participants more directly, which is the goal of Experiment 2. Convergence in production for Experiment 2 is assessed using the same Word Naming Game paradigm as Experiment 1, with some minor adjustments. For one, in order to directly compare expectation-driven convergence to input-driven convergence, a third between-subjects condition is introduced in Experiment 2, in which the same Southern talker used in the Southern (/aɪ/–absent) Condition instead *does* produce glide-weakened /aɪ/ throughout the course of the experiment. This is referred to as the Southern /aɪ/-Present Condition, and is used to assess input-driven convergence. The other two conditions (Midland (control) and Southern /aɪ/-Absent are identical to those in Experiment 1. Perceptual shifts are tested with a between-subjects exposure-test perceptual learning paradigm, specifically using an auditory lexical decision design in which participants judge sound clips as real "Words" or "Non-Words." Target items are glide-weakened /aɪ/ words spoken by the Southern model talker, that may be confused for non-words with their glide-weakened productions (e.g., "bribe" that sounds like "brahb"). This task is administered directly after the Exposure Phase of the Word Naming Game Task, such that the exposure phase serves as the exposure stimuli for both the production and perception task. It is predicted that participants who have recently heard a Southern talker (in the /aɪ/–absent or /aɪ/–present conditions) will have higher "word" identification rates for target items than those who heard the Midland talker. Finally, individual differences are measured with a series of surveys administered at the end of the experiment. Individual difference measures that tap into domain-general capabilities (e.g., Openness, Attention Switching, or Behavior Inhibition), have been shown to correlate with linguistic behavior, including linguistic convergence. Comparing whether the same individual difference measures predict both expectation-driven and input-driven shifts in perception or production will offer some insights into the relationship between these two behaviors. The measures utilized here are the Big Five Inventory (BFI) (John and Kentle, 1991), The Autism-Spectrum Quotient (AQ) (Baron-Cohen et al., 2001), and the Marlowe-Crowne Social Desirability Scale (Crowne and Marlowe, 1960). Attitudes toward the South and towards the model talker are also assessed during these follow-up surveys. The overall design is shown in Figure 2.2.

2.3 Experiment 3

The goal of Experiment 3 is to understand the potential mechanisms that may be responsible for the expectation-driven convergence observed in Experiment 1. Specifically, Experiment 3 works to broadly disentangle socially-rooted "top-down" processes from structurally-rooted "bottom-up" processes. A socially-rooted, top-down mechanism would crucially rely on identifying the talker as a member of the social category "Southern." Activating this broad category label of "Southern-ness" may then trigger phonetic features associated with this category, including glide-weakening of /aɪ/. A structurally-rooted, bottom-up explanation is also possible, in which speakers rely on knowledge of variant co-occurrence, regardless of the social label assigned to the talker. Observing the acoustic properties of other Southern-shifted vowels would activate glide-weakened /aɪ/ by association. Another way of thinking about this is, is knowing someone is from the South enough to elicit convergence, or is it necessary to recognize that they are also Southern-shifted?

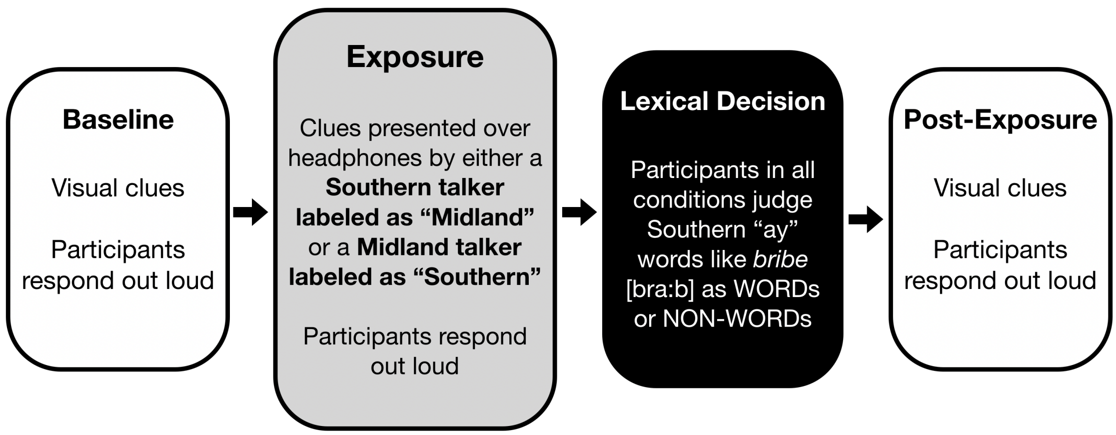

Figure 2.3: Experiment 3 Design

Top-down and bottom-up effects are manipulated with a dialect-label mismatch manipulation of the general Word Naming Game design. Explicit talker labels provide top-down social cues, while the different accents of the model talkers provide bottom-up acoustic cues. However, it is non-trivial to isolate these different influences. Regardless of the explicit top-down information provided about a talkers' accent, whenever a listener actually communicates with the talker, they will be exposed to bottom-up acoustic cues. Likewise, when a listeners hears an accented voice, there is nothing to stop them from positing social category membership of the talker based on various qualities of their speech. Considering both types of information are available to listeners, it is impossible to say which information source actually activates glide-weakened /aɪ/. In an attempt to isolate top-down and bottom-up information as much as possible, each of the model talkers gives conflicting top-down and bottom-up cues, where only one of these sources of information would be consistent with shifts toward Southern speech. In the "Voice-Triggered" Condition, participants

	Experiment 1	
Southerners	Midland Condition: 25 Southern Condition: 27	
Non-Southerners	Midland Condition: 29 Southern Condition: 31	
	Experiment 2 - Production *	**Experiment 2 - Perception ***
Southerners	Midland Condition: 29 Southern Condition (/aɪ/ absent): 33 Southern Condition (/aɪ/ present): 29	Midland Condition: 34 Southern Condition (/aɪ/ absent): 38 Southern Condition (/aɪ/ present): 35
Non-Southerners	Midland Condition: 26 Southern Condition (/aɪ/ absent): 26 Southern Condition (/aɪ/ present): 25	Midland Condition: 30 Southern Condition (/aɪ/ absent): 30 Southern Condition (/aɪ/ present): 36
	Experiment 3 - Production **	**Experiment 3 - Perception ****
Southerners	Midland Voice/Southern Label: 26 Southern Voice/Midland Label: 25	Midland Voice/Southern Label: 37 Southern Voice/Midland Label: 38
Non-Southerners	Midland Voice/Southern Label: 32 Southern Voice/Midland Label: 32	Midland Voice/Southern Label: 32 Southern Voice/Midland Label: 32

* 159 participants have combined perception-production results for Experiment 2. ** 112 participants have combined perception-production data for Experiment 3.

Table 2.1: Participant breakdown across the three Experiments.

hear a Southern-accented talker but are told the talker is from the Midwest and has a "Midwest accent," so only bottom-up information would provide cues to activate glide-weakened /aɪ/. In the "Label-Triggered" Condition, participants hear a Midland-accented talker but are told the talker is from the South and has a Southern accent; in this case, only top-down information would provide cues to activate glide-weakened /aɪ/. Based on previous literature suggesting that both top-down (Niedzielski, 1999; Koops et al., 2008; Hay and Drager, 2010) and bottom-up (Theodore and Miller, 2010) information triggers expectations, I predict that both types of cues can elicit glide-weakened /aɪ/, though the effect may be somewhat weaker due to the conflicting information provided by the two different sources of information.

2.4 Southern Speech

The dissertation focuses on Southern speech as a test-case in probing expectation-driven convergence. Specifically, each experiment tests expectations about "Southernness" and glide-weakened

/aɪ/. The Southern U.S. is perhaps the most distinct dialect region in the U.S., outlined in Figure 2.4, below, and glide-weakening of /aɪ/[1] is a particularly salient and stereotypical feature of Southern English (e.g., Hall, 1942; Wolfram and Christian, 1976; Reed, 2014). As Reed (2016) notes, "American English speakers are aware of monophthongization and its status as a regional and sub-regional linguistic caricature...Virtually every popular depiction of Southern and Appalachian speech displays monophthongal /aɪ/ as a noteworthy feature..." He goes on to suggest that this feature would be considered a *stereotype* under Labov's 1972 classic categorizations of sociolinguistic variables. Further, Torbert (2010) found that /aɪ/ glide-weakening is a salient perceptual cue indexing "Southernness," though this was not the case for another Southern dialect feature, fronting of /o/. Though there has not been much working comparing the salience of Southern /aɪ/ to that of other Southern features, it is generally agreed that /aɪ/ glide-weakening is among the most salient. As Labov (2010) notes, "The most generally recognized feature of Southern speech is the monophthongization of /aɪ/" (p. 55).

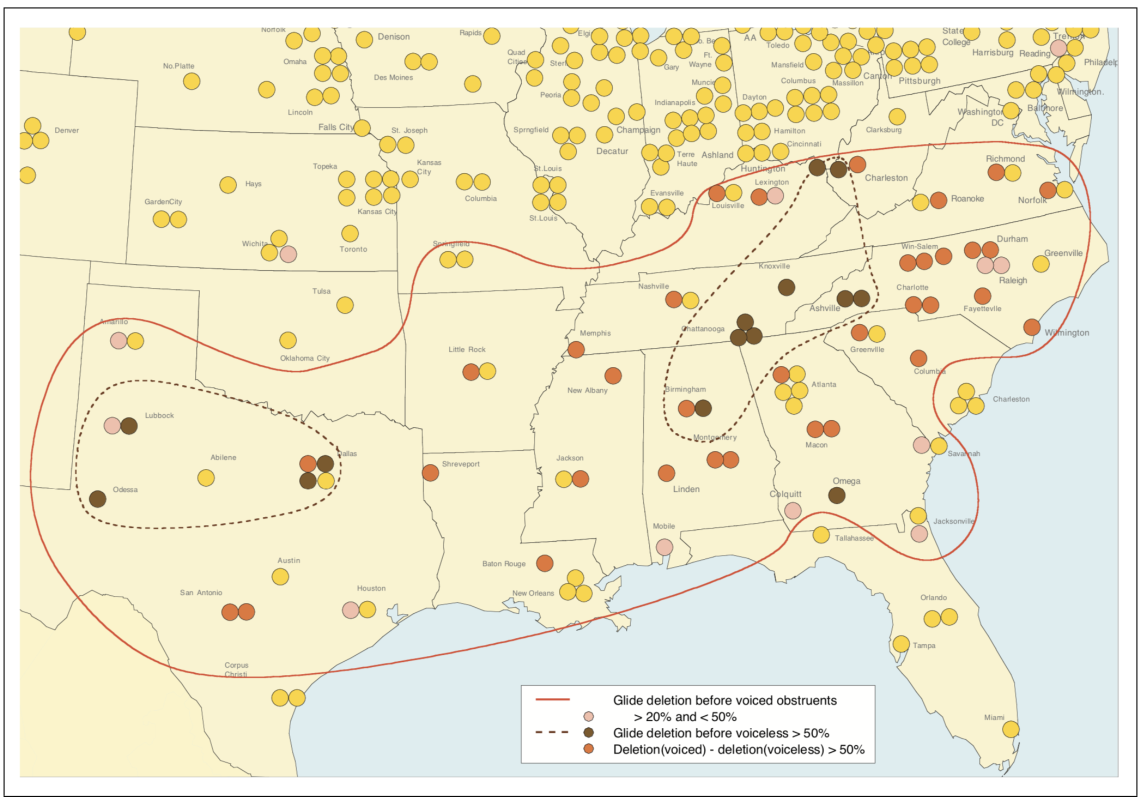

Figure 2.4: Southern isogloss from Labov et al. (2006)

[1]The phenomenon referred to as "glide-weakening" is also commonly called "glide reduction," "glide deletion," "ungliding," or "monophthongization."

Weakening of the /aɪ/ glide is often described as the triggering even of the "Southern Shift," a vocalic chain shift affecting practically the entire vowel space. This shift is considered a "pull" chain shift (as opposed to a "push" chain shift) because the vowels move in order to fill empty spaces caused by previous vocalic movement. The shift is thought to have begun during the mid-to-late 1800s (Bailey, 1997) and is still progressing in many rural regions of the South, though it is retreating in more urban areas (Labov et al., 2006; Dodsworth and Kohn, 2012; Dodsworth, 2014). The shift affecting the front vowels is often described as comprising three stages, beginning in Stage 1 with the monophthongization or glide-weakening of /aɪ/. The phonological constraints and social stratification of this shift vary throughout the South: In more inland regions of the South (as shown in Figure 2.4, and among working class speakers, glide-weakening is more likely to occur in all phonological environments, including before voiceless segments. However, in the majority of the South, glide-weakening only occurs in coda position or preceding voiced segments. Regardless, when /aɪ/ glide-weakening occurs in the South, /aɪ/ shifts slightly forward and enters the system of long, ingliding vowels and exits the system of upgliding vowels (Labov, 1994). This triggers Stage 2, which begins with the nucleus of /e/ shifting downward along a non-peripheral path, essentially taking the place of /aɪ/ (Labov et al., 2006). /e/, once on a non-peripheral path, follows Labov's (1994) Principle II of chain shifting: *Lax nuclei fall along a non-peripheral track.* This movement of /e/ allows room for the nucleus of /ɛ/ to tense, then follow Principle I of chain shifting: *Tense nuclei raise along a peripheral track.* The nuclei of /e/ and /ɛ/ are typically discussed in terms of "reversing" their positions in relation to one another. Stage 3 extends this reversal to the higher /i/ and /ɪ/ vowels, which follow a parallel shift in some (but not all) regions of the South, particularly in the inland South. While the Southern Shift of the front vowels is often simplified and referred to as a "reversal" of the front tense and lax pairs, the tense pairs actually fall further than this, so that /e/ approaches /aɪ/ and /i/ approaches /e/. Short-/a/ also follows along with the other lax vowels by tensing then rising along a peripheral path. This movement is illustrated in Figure 2.5.

The back vowels also shift in the South, though this is not typically considered to be part of the chain shift just described. As Labov et al. (1972) note: "It is not at all obvious that a chain shift is involved in this situation, since there seems to be no back upgliding vowel which moves up behind /u/ and /o/ to assume their positions and which might have been held back by /u/." Instead, the back upgliding vowels front in parallel, following Labov's (1994) Principle III of chain

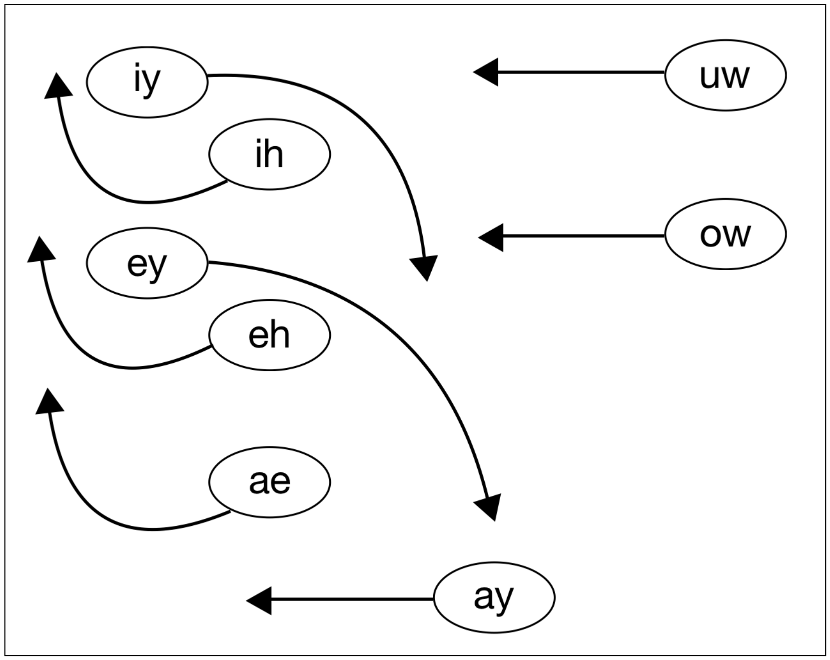

Figure 2.5: The Southern Shift

shifting: *back vowels move to the front.* Note though that the fronting of /u/ is in some cases so extreme that it is realized phonetically directly behind /i/ (Labov et al., 1972, 2006). However, the fronting of these back vowels is not unique to Southern phonology. In fact, fronting of /u/ occurs to some extent in most of North America and while fronting is common in the Midlands as well, /o/ fronting is most advanced in the South. These are not the only features of Southern English. For instance, velar nasal fronting, the *pin-pen* merger and vowel "breaking" (i.e., the "Southern Drawl") are associated with the South. Further, another chain shift, the "back upglide shift," which involves fronting of /aw/ (as in *house*) and upgliding of /oh/ (as in *caught*) occurs in the South.

While the experiments presented in the dissertation focus primarily on /aɪ/ glide-weakening, it is important to recognize this phenomenon as part of the larger Southern-shift and to note the other Southern-shifted forms that participants *do* observe from the Southern model talker in the subsequent experiments when considering the potential acoustic cues triggering expectations for /aɪ/ glide weakening. The Southern U.S. dialect—and /aɪ/ glide-weakening specifically—was chosen for this study because of its distinctiveness and social salience. In an initial investigation of the role of social expectations in convergence, the best starting place is with a sociolinguistic feature that is most likely to generate expectations.

Chapter 3

Experiment 1: Establishing Expectation-Driven Convergence

This chapter reports the results of Experiment 1, which investigates whether expectation-driven convergence occurs in a controlled experimental setting. Specifically, I investigate whether participants converge toward a regional variant they might reasonably expect, but do not directly observe, from a model talker. The variant of interest is glide-weakened /aɪ/, a salient feature of Southern U.S. English. To investigate this question, I developed a novel Word-Naming Game paradigm for targeted speech elicitation, in which clues are used to elicit /aɪ/ tokens from participants before and after exposure to a Southern or Midland (control) model talker. The game-like nature of the task is designed to promote feelings of interaction with the model talker, thereby facilitating convergence and reducing potentially inhibitive effects of speech monitoring.

In addition to testing whether expectation-driven convergence occurs, the experiment also has several exploratory aims centered on the nature of expectation-driven convergence that will lay the groundwork for the subsequent experiments presented in Chapters 4 and 5. For one, the experiment investigates some of the factors that may facilitate expectation-driven convergence, including dialect experience, cognitive/personality traits, and social attitudes. Understanding how both situational variables and stable individual traits impact the likelihood and degree to which an individual will converge may hint at the cognitive processes utilized in convergence behaviors generally, which is the main focus of Experiment 2 (Chapter 4). Additionally, the experiment compares convergence toward glide-weakened /aɪ/ (which is only *expected*) and other vowels that participants *do* observe within the context of the experiment, which informs the more direct comparison of input-driven and expectation-driven convergence in Chapter 4. The role of dialect background is also examined by comparing the behavior of participants from within and outside of the South, as well as exploring how participants' own baseline linguistic behavior influences their tendency to converge.

There are several specific research questions to be addressed in this chapter:

1. Do participants converge toward a variant they expect—but do not observe—from a model talker?

2. If so, how long does convergence last post-exposure?

3. How does sociolinguistic experience and/or dialect background mediate convergence?

4. Do individuals conform to the aggregate group pattern? If not which individual factors (both stable and situational) may predict convergence rates?

5. Do individuals converge toward vowels they *do* observe from the model talker in the context of the experiment, and how does this compare to converge toward the unobserved variant /aɪ/?

Section 3.1 first reviews the relevant literature on the existence of expectation-driven convergence, and Section 3.2 outlines the methods used in Experiment 1, which are the same methods used for the Word Naming Game tasks in subsequent experiments. Section 3.4 reports the findings of Experiment 1, which establish expectation-driven convergence as an empirically observable linguistic phenomenon, and suggest that degree of convergence is predicted by both dialect distance and motivations to appear socially desirable. Section 3.5 then summarizes and interprets these results in light of existing findings and outlines remaining questions.

3.1 Background

Here, I review the relevant literature informing the research questions and predictions of Experiment 1. Since the main goal of this chapter is to experimentally test whether expectation-driven convergence occurs, I focus here on reviewing previous reports of expectation-driven convergence, as well as experimental evidence for behaviors that might also be thought of as related to expectation-driven convergence.

3.1.1 Evidence for expectation-driven convergence

Previous reports of expectation-driven convergence have been primarily anecdotal. For instance, in Fasold's (1972) work on African American English in Washington D.C., he observed that speakers produced more African American English (AAE) features such as [d]-deletion when conversing with an African American interviewer compared to a White interviewer, though most of the interviewers were middle-class women who spoke Mainstream American English (MAE) rather than AAE. Similarly, Bell (2001) reported that an Anglo interviewer in New Zealand frequently used the *eh* tag—a feature stereotypically associated with male Maori speech—when conversing with a male Maori interviewee but not when conversing with an Anglo interviewee. Importantly, though, the Maori interviewee never actually used this feature, suggesting that the interviewer's behavior was a form of convergence based on expectations rather than observed linguistic behavior. Auer and Hinskens (2005) use this anecdote as support for an "Identity-Projection Model" of convergence, which suggests that instead of converging toward "observable behaviour of the recipient ... speakers converge to a stereotype of the 'model' receiver, not the actual partner in direct communication" (p. 341). That is, speakers may change the way they speak to match how they believe their conversational partner speaks, regardless of whether their conversational partner actually speaks that way. The idea of convergence triggered by the speaker rather than the speech itself can be traced further back to Thakerar et al. (1982), who terms accommodation which "responds to what the speaker mistakenly assumes will be the addressee's speech on the basis of the addressee's non-speech attributes" to be "subjective accommodation" (Bell, 1984, p. 168).

More recently, expectation-driven convergence has been observed in an experimental study of computer-mediated communication in an artificial "alien" language. Wade and Roberts (2020) found that participants who were assigned to play one of two alien species who exhibited dialectal variation converged by producing the form associated with the species of their partner, both *before* observing their partners' actual language usage and *despite* observing contradictory behavior from their partner who used a form not expected for their species. This is perhaps the first laboratory study providing evidence for expectation-driven convergence, and these results suggest that expectation-driven convergence may be a promising area for probing the nature of sociolinguistic expectations. It is important, moving forward in this line of work, to test the generalizability of these findings by examining how similarly individuals behave when conversing aloud in natural language.

In addition to the findings reported above, other types of reported linguistic behavior might reasonably be driven by the same mechanisms underlying expectation-driven convergence. One of these behaviors is holistic shifts toward entire linguistic systems as opposed to specific individual linguistic targets. There has been some recent work showing that expectations about one component of a talker's linguistic system can be cued by other aspects of the system. For instance, studies on perceptual learning have shown that listeners exposed to a subset of a previously unheard vowel chain shift were able to fill in gaps and generalize to phonemes that were not a part of training (e.g., Weatherholtz, 2015). Theodore and Miller (2010) also demonstrated that listeners can make talker-specific predictions about one aspect of a talker's linguistic system after observing a different aspect. They found that observations of VOT for /p/ for and individual talker informed listener predictions of VOT for /k/. Work on convergence has similarly reported more holistic linguistic targets. For instance, as noted in Chapter 1, Zellou et al. (2017) suggested that listeners are influenced by a more holistic model of a talker rather than individual linguistic tokens, based on observations that participants converged toward the the average nasalization observed from a talker rather than most recent tokens. These behaviors all might reasonably be considered instances of expectation-driven convergence in that participants aren't shifting (in perception or production) toward isolated tokens they've observed, but rather to values they *expect* based on linguistic features observed from the same talker.

Style-shifting is another behavior that may, in some instances, actually be a result of expectation-driven convergence. Bell's (1984) Audience Design Model of style-shifting, for instance, allows for speech to be influenced by not only one's primary addressee, but also by others who may be involved in the communicative situation (auditors, overhearers, eavesdroppers), as well as other non-personal variables such as topic and setting. The Audience Design model assumes that, when people speak, they take into account primarily those who would be listening to that speech and design their speech accordingly. While there are a number of ways in which an individual may shift their speech when taking those around them into account, perhaps the most obvious is to converge *toward* the linguistic behaviors expected from these individuals. This is consistent with our definition of expectation-driven convergence as a shift in speech toward an addressee that may be triggered by non-speech factors. Other findings in the realm of sociolinguistic style-shifting may also be considered instances of expectation-driven convergence. For instance, Rickford and McNair-Knox (1994) found that an African American teenager "Foxy Boxton" used more AAVE features when conversing with African American interviewers. Still, the relationship between style-shifting and convergence remains largely unclear. While style-shifting is often (though not always) assumed to involve shifting along a single spectrum of formality within a given speakers' repertoire and convergence requires an explicit interlocutor to converge toward, there is much overlap in the types of behavior that could be considered style-shifting or convergence. Evidence that speakers shift their speech to include particular variants that they do not directly observe in that moment, makes expectation-driven convergence a likely behavior for speakers to exhibit as well.

3.2 Methods

Linguistic convergence is a well-established phenomenon that has been observed in both naturalistic, dyadic settings as well as in laboratory experiments. Eliciting convergence in the lab has been quite successful and laboratory findings have been consistent with observations of convergence in dyadic conversation (Pardo et al., 2018). Moreover, laboratory experiments are useful in that they allow for greater control over the tokens elicited, as well as some control over social factors. However, typical methods of investigating convergence in the lab don't work for expectation-driven convergence. Convergence is usually investigated experimentally using shadowing tasks, in which participants read words aloud as a baseline speech measure, and then repeat words after a model talker as a shadowed speech measure. Baseline and shadowed speech are then compared to determine whether participants shift toward the model talker when shadowing. Expectation-driven convergence cannot be investigated using a typical shadowing design where words are repeated after a model talker because, in order for expectation-driven convergence to occur, participants have to shift their speech toward a target that they crucially never hear from a model talker. Other speech elicitation methods such as cooperative map tasks, Diapix (find the difference) tasks, or naturalistic dyadic conversations are also sometimes used to investigate convergence. Such tasks may get around the issue of needing to repeat after a model talker, though they pose a new set of difficulties for investigating expectation-driven convergence. For one, eliciting a sufficient number of /aɪ/ tokens is a nontrivial task. More importantly, it would be extremely difficult to ensure one member of the dyad avoids any usage of the variant of interest. Any methodological paradigm designed for investigating expectation-driven convergence must therefore avoid these complications, while allowing for sufficient exposure to a model talker, ideally enhancing the likelihood of convergence by facilitating feelings of real interaction with the model talker.

While objective acoustic measures are often used to determine whether convergence has occurred (e.g., comparing VOT or F0 in the baseline phase to that in the shadowing phase), it is also frequently the case that more subjective measures such as AXB experiments are conducted to determine whether more holistic convergence has occurred. In an AXB task, a new group of participants hears the model talker's production of a word, then judge whether a participant's baseline or shadowed production sounds more similar to the model talker's. Such a measure of holistic convergence, however, is not easily adaptable for a study of expectation-driven convergence since participants' production cannot be directly compared to anything the model talker produced in the course of the experiment. Therefore, only single-dimension acoustic measures can be used.

No existing speech elicitation paradigm seemed to sufficiently meet the criteria outlined above. In order to investigate expectation-driven convergence, I therefore created a novel experimental paradigm, the Word Naming Game. The premise of the task is that participants read (baseline) or hear (shadowing) clues describing a target word, which they must then guess aloud. Their responses to orthographically–presented clues are then compared to their responses to auditorily–presented clues to determine whether convergence has occurred. Importantly, this task allows for elicitation of specific tokens without the model talker producing them, and provides sufficient exposure to the model talker's voice. The procedure for this task is described in more detail below, and this same paradigm is used for all three experiments presented in the dissertation.

3.2.1 Participants

120 participants were recruited to participate in Experiment 1. Of these 120 participants, 105 were recruited through Prolific and completed the experiment online for $10 USD. Fifteen participants were recruited through the University of Pennsylvania subject pool and participated in the Language Variation and Cognition Lab in the University of Pennsylvania's Linguistics Department then

received course credit for their participation. Due to recording failures or excessive (>20) mistakes in the task, 8 participants were excluded, and the data from the remaining 112 participants will be presented here.

Of these 112 participants, 56 self-identified as female, 53 as male, 2 as non-binary, and 1 did not disclose. Since many participants were collected from the online Prolific platform, the data presented here includes a greater mix of ages, education levels, and racial backgrounds than we typically see in experimental work on college students. Regarding maximum educational attainment, 5 reported less than a high school diploma, 20 reported a high school diploma, 2 reported technical training or certification, 38 reported some college, 10 reported an Associate's degree, 28 reported a Bachelor's degree, and 8 reported a graduate or advanced degree. Due to an error in recording racial/ethnic self-identification on the first half of surveys administered, we report this data for only 62 participants: 37 identified as White, 14 as Black or African American, 4 as Asian or Pacific Islander, 4 as Hispanic or Latinx, and 2 as other. Mean age is 29.4 (Range 18–60). Participants were all native speakers of American English and reported no speech or hearing impairments.

Participants were labeled as "Southern" or "Non-Southern," depending on their dialect background. Participants who spent the majority of their school-aged years (ages 5-18) within the outlined area in Figure 2.4, depicting the isogloss for /aɪ/ glide-weakening were labeled as "Southern." Fifty-two participants included in this study were classified as "Southern." The remaining 60 participants spent the majority of their school-aged years outside of the U.S. South and were classified as Non-Southern. In general, the Southern participants produced only slightly more Southern vowels than Non-Southern participants, and the vast majority of Southern participants were not recognizably Southern-shifted. As Figure 3.1 shows, both groups' /i/, /ɑ/, and /o/ categories are roughly the same, Southern participants' /ɪ/ and /æ/ are only slightly higher and fronter (i.e., more Southern-shifted). Further, /u/ is largely fronted for both groups, though this fronting extends somewhat further for Southerners. /ɛ/ and /e/ show largely overlapping categories for Southerners and a general distinction for Non-Southerners. The /aɪ/ vowel, with mean values for Southerners (triangle) and Non-Southerners (circle), indicated on the graph, is also produced similarly across groups. Figure 3.2 focuses on just the /aɪ/ vowel, with points representing mean values for each speaker. As we can see, there is little overlap between speaker means of the /aɪ/ glide and nucleus, suggesting a mostly diphthongal vowel for most speakers.

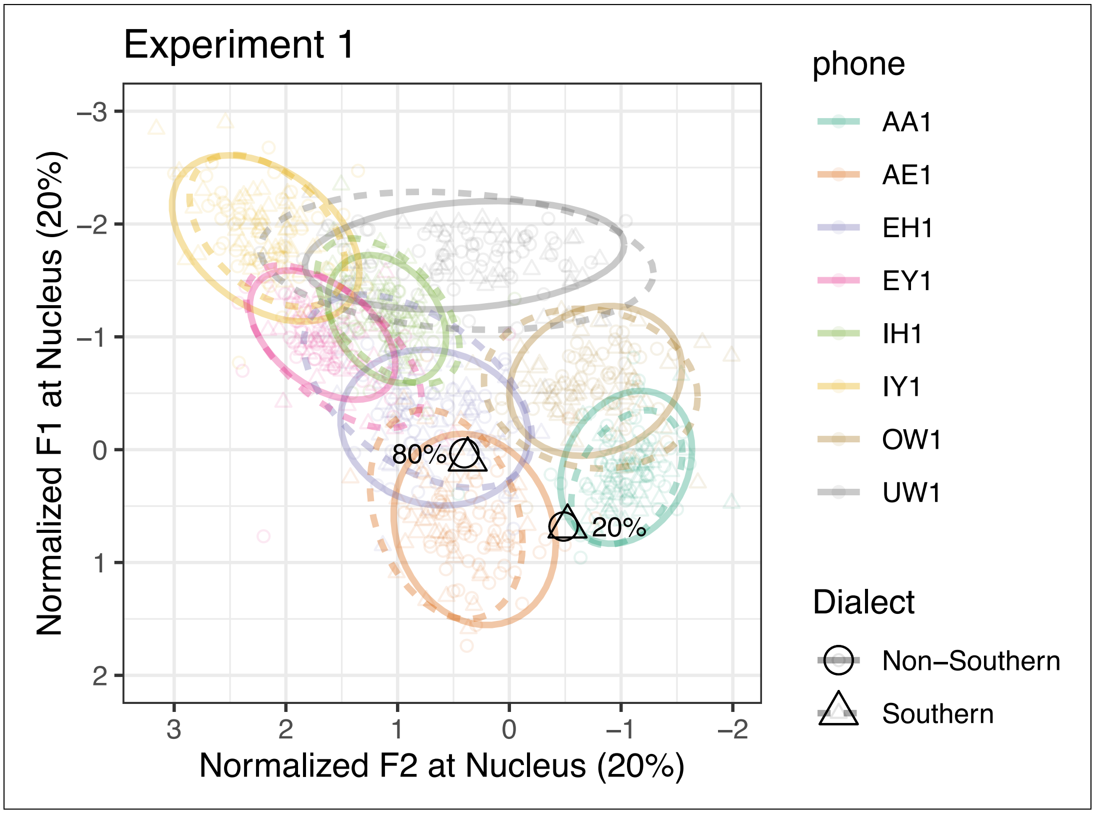

Figure 3.1: Southerner vs. Non-Southerner vowels. Each point represent participant means, and linetype indicates dialect group. Shapes indicate Southerners' (triangle) and Non-Southerners' (circle) productions of the /aɪ/ vowel at the nucleus (20%) and glide (80%).

3.2.2 Procedure

The experiment was programmed in Ibex (A. Drummond, www.spellout.net/latest_ibex_manual.pdf) using the PennController (Zehr and Schwarz, 2018). The task was completed in an internet browser, where Participants' voices were recorded through their computer microphone. After giving informed consent, participants tested their microphone and headphones. Participants then completed a brief demographic questionnaire, then completed the three phases of the Word Naming Game, followed by a follow-up survey consisting of (1) Demographic questions (2) Attitudinal questions toward the model talker (3) Attitudinal questions about the South (3) The Big Five Inventory (John and Kentle, 1991) (4) The Autism-Spectrum Quotient (Baron-Cohen et al., 2001) and (5) the Marlowe-Crowne Social Desirability Scale (Crowne and Marlowe, 1960). These different components are discussed in detail below.

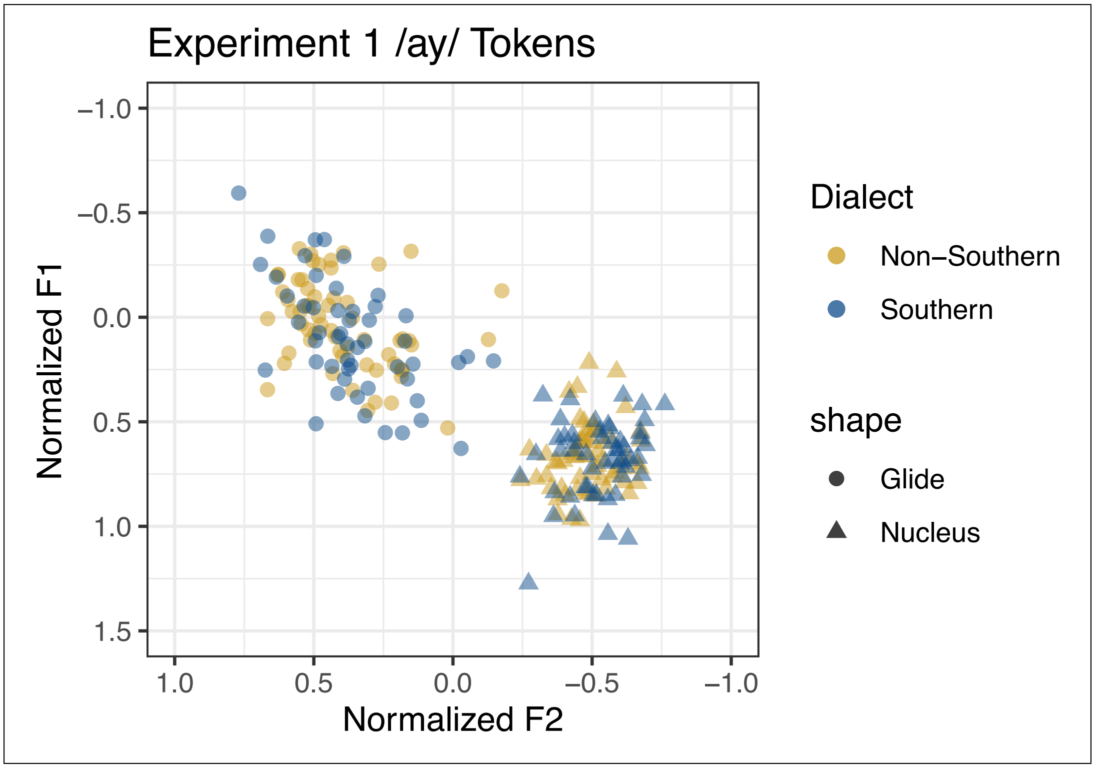

Figure 3.2: Individual speaker means for the /aɪ/ vowel, with triangles indicating the nucleus (20%) and circles indicating the glide (80%).

Demographic Questionnaire

After providing informed consent to participate in the experiment, participants completed a brief demographic questionnaire consisting of the following questions:

- Select your age.
- What gender do you identify as?
- Which race/ethnicity do you most identify with?
- What is the highest level of education you have obtained?
- Please enter all cities you have lived in and the ages you lived there. Example: Cleveland OH (birth-5), Raleigh NC (5-20).

The final question was particularly important for labeling participants as “Southern” or “Non-Southern” based on where they spent the majority of their school-aged years.

Word Naming Task

The bulk of the experiment consisted of the Word-Naming Game task, described briefly above. In this task, "clues" were presented to participants in written or audio format, depending on the experimental phase, and participants named aloud the word described by the clues into their computer microphone, using the carrier phrase, "The word is ____." For instance, a participant may be given the clue, "The saying goes, if at first you don't succeed, do this three letter T-word again," and would respond aloud, "The word is 'try.'" To record their response, participants pressed a "Record" button on the screen, and the screen indicated that they were recording. When they were finished recording the response, participants pressed the "Next" button to continue to the next clue, and the screen indicated that they were no longer recording. Hints were also given for each item to encourage correct responses and prevent data loss. Hints consisted of a number of blanks indicating the number of letters in the word, with 1–3 letters filled in. An example of what the screen looked like during a trial is shown in Figure 3.3.

The experiment consisted of three phases. The BASELINE PHASE (Phase 1) elicited participants' baseline vowel measurements prior to any auditory exposure. In this phase, clues were presented orthographically on the screen and participants read the clues silently to themselves before giving their responses aloud. The EXPOSURE PHASE (Phase 2) served as the experimental phase. Clues were presented in audio format over headphones, though the task was the same. Clues in this phase were given aloud by either a Southern talker or a Midland talker (control), depending on the condition to which the participant was randomly assigned. I take the acoustic measurements of participant's vowel productions in the Exposure Phase, compared to those in the Baseline Phase, as a measure of convergence. Note, however, that this measure could capture other shifts in vowel production across the experiment due to other factors. For instance, since this experiment targets /aɪ/ glide-weakening, a reduction process, the result of general fatigue as the experiment progresses (i.e., more reduced vowels) may be confused for convergence toward a more monophthongal production of /aɪ/. For this reason, the Midland Condition is included as a control. In the final, POST-EXPOSURE PHASE, clues were again presented in written format. A Post-Exposure Phase was included in order to investigate the longevity of convergence effects, but can also be used to tease apart effects of fatigue from production shifts stemming from the experimental manipulation.

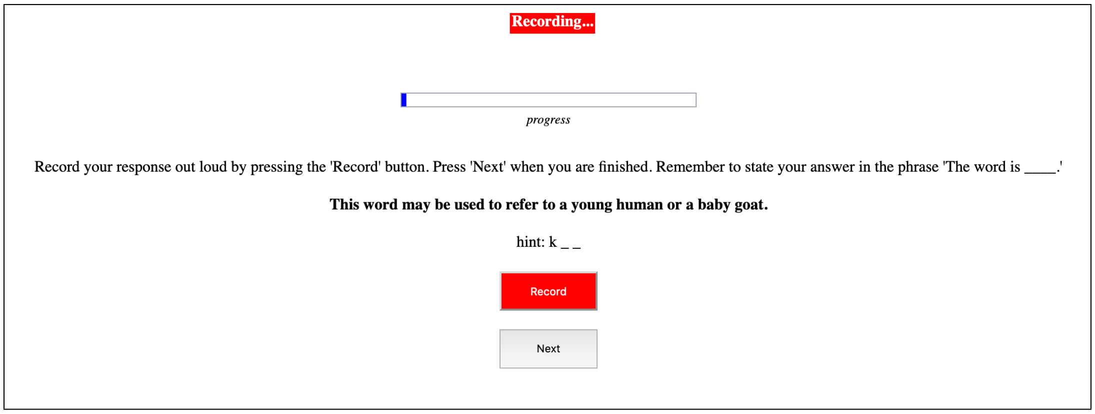

Figure 3.3: Screen shot of what a trial in baseline/post-exposure phase, where clues are presented on-screen.

At the beginning of the experiment, directions were presented orthographically on the screen, as follows:

> *You will be playing a game today. I will try to get you to guess a single word by giving clues. The clues can be very specific but cannot contain the word itself. You will guess the word being described by naming it out loud. To record your response press the 'Record' button. When you are finished, click 'Next' to continue onto the next clue, and your response will be saved automatically.*
>
> *An important note: When you give your response, you must say it in the phrase 'The word is ___.'*
>
> *For instance, if the clue is 'This is a household pet that often meows', you would respond by saying 'The word is cat.'*

Before the auditory presentation of clues in Phase 2, the directions were given again over the headphones so that participants could familiarize themselves with the model talker. The auditory instructions were somewhat different from the earlier orthographically–presented instructions, partially to omit any words containing the /aɪ/ vowel. The model talker's script is as follows:

> We are attempting to determine whether participants are better at guessing words when the descriptions are presented in written vs. spoken format. For the next section, the clues will be given to you out loud. The instructions are the same. Each correct answer will be a single word. You should embed it in the phrase "The word is BLANK." For

> example, if the clue is "This household pet often meows," you would respond "The word is CAT." When you are ready to begin press the space bar on your keyboard. Please do not begin each response until the clue has finished. Let's begin!

Crucially, the /aɪ/ vowel was never used in any of the clues or in the directions. Any incorrect participant responses were omitted from data analysis. In order to minimize data loss due to incorrect responses, participants were provided with a hint at the bottom of the screen for all clues, which contained spaces indicating the number of letters in a word, with some letters filled in (e.g., *t _ _ r*).

Post-task survey

After the Word Naming Task, participants completed a post-task survey, which consisted of six components. The first component included general demographic information and questions regarding what participants believed the experiment to be about and perceived regional identity of the talker. The second component asked participants to rate the model talker they heard on 12 personal attributes, using a Likert scale ranging from 0-100 (e.g., "The talker sounded kind"). The third component listed 9 statements about the South and asked participants to determine the extent to which they agreed with the statement, using a Likert scale ranging from 0-100 (e.g., "I am familiar with the way Southerners speak"). The statements that participants rated for these sections are provided in Appendix A. The fourth component was the Big Five Personality Inventory (John and Kentle, 1991), a self-assessed psychometric survey that measures an individual on five personality dimensions: extraversion vs. introversion, agreeableness vs. antagonism, conscientiousness vs. lack of direction, neuroticism vs. emotional stability, and openness vs. closedness to experience. The fifth component included was the Autism Spectrum Quotient (AQ) questionnaire (Baron-Cohen et al., 2001), a self-reported measure of "Autistic-like" traits, that has been administered to a neurotypical population as a measure of cognitive processing style (Yu, 2010; Yu et al., 2013). Sub-categories measured by the AQ include Attention Switching, Social Skills, Imagination, Communication, and Attention to Detail. The final measure was the Marlowe-Crowne Social Desirability Scale, which measures the extent to which an individual misrepresents themselves to earn the approval of others (Crowne and Marlowe, 1960). This self-reported measure consists of questions that are unlikely to be entirely true, but that are socially desirable (e.g., "I have never been irked when people express ideas different from my own"), and is often used to determine how truthful an individual will be on other self-assessment surveys. These various cognitive-style/personality measures were included for two main reasons. The first is to test whether the individual differences measures previously shown to predict degree of phonetic convergence are the same ones that predict degree of convergence in the present study. The second is to determine whether similar individual difference measures predict convergence for both input-driven and expectation driven-convergence, which will shed light on whether these behaviors are stem from similar cognitive processes.

3.2.3 Stimuli

Elicited tokens consisted of 90 /aɪ/ words in word-final or pre-voiced consonant contexts, as this is the environment in which /aɪ/ glide-weakening reliably occurs in the U.S. South. 90 non-/aɪ/ filler words were also elicited. A complete list of elicited words is provided in Appendix B. These tokens were elicited in three sets consisting of 30 /aɪ/ words and 30 fillers each, with one set being elicited in each phase, such that no item was repeated within the experiment. The phase in which each set was elicited was counterbalanced across participants, and individual items within each set were randomized for each participant. The three sets of elicited words were balanced for place of articulation of adjacent segments, as well as for lexical frequency, determined using the SUBTLEX corpus (Brysbaert and New, 2009) Log10CD measure. For each vowel class, frequency of words across sets had the same mean and standard deviation. The three lists of target items with their frequency measures are included in Appendix B.

Clues were 1-2 sentences long and gave participants some evidence for the dialect of the talker. A complete list of clues used to elicit each item is given in Appendix B. Crucially, the /aɪ/ vowel was never included in any of the clues. When clues were presented auditorily, they were read by either a Midland or Southern young, white, female talker. The Southern talker is from Hurley, Mississippi and produces typical Southern dialect features such as reversal of the front tense and lax vowel nuclei, fronting of back vowels, and PIN-PEN merger. The Midland speaker is from Youngstown, Ohio and was described by participants as "normal" or "neutral" sounding.

3.2.4 Analysis

Recordings were forced-aligned using the Penn Phonetics Lab Forced Aligner (p2fa) (Yuan and Liberman, 2008). Since shifts in convergence are often relatively small and may be difficult to detect in noisy or highly variable speech, additional steps were taken to ensure a high degree of precision and accuracy in vowel measurements. Each TextGrid was blindly hand-corrected[1] and formant tracking settings in Praat were adjusted for each speaker, and each token as necessary, with defaults of 5 number of formants and 5500 max formant height for females and 5000 for males. A script was run in Praat (Boersma and Weenink, 2002) to automatically measure the first three formants at 7 time points throughout the course of each vowel: 10%, 20%, 30%, 50%, 70%, 80%, and 90%. Formant measurements were then normalized using the Lobanov method (Lobanov, 1971). Any incorrect responses given were omitted from the analysis, and all data visualization and statistical modeling was done in R (R Core Team, 2015). The primary measure used in the following analysis will be the height of the glide, F1 at 80% into the vowel; this decision was made based on the results of the pilot experiment, described in Section 3.3 below. Outliers were excluded from the following analysis on a by-item and by-participant basis, with observations greater than 3 standard deviations from the mean being omitted. This leaves a total of 9,491 /aɪ/ tokens (or roughly 84 per person and 28 per-person per experiment phase) in the analysis.

3.3 Pilot Experiment

The experimental design was piloted before running the experiment as described above and results are reported in Wade (2017a). The pilot experiment followed the same general design but with several small changes: the pilot (1) used two male model talkers in their 30s rather than the two female talkers described above (2) was completed in person with the experimenter present rather than remotely, online (3) recruited Southerners from North Carolina State University and Non-Southerners from the University of Pennsylvania, (4) included some items that participants failed to guess which were then omitted or replaced with other words in Experiment 1, and (5) did not contain "hints," which were added to Experiment 1 to avoid data loss from participants not guessing

[1] "Blindly" hand-corrected refers to the fact that TextGrid editors did not know which condition a participant was in, or what experimental phase a soundfile came from when editing.

the correct items. In this pilot version, the Midland talker was from Northeast Ohio and the Southern Talker was from Western North Carolina and exhibited typical Southern-shifted features. All other procedures were the same as in Experiment 1, as described above.

In total, 84 participants from either the University of Pennsylvania or North Carolina State University were recruited for the pilot study. Participants were labeled as "Southern" or "Non-Southern," depending on their dialect background. Participants who grew up within the outlined area in Figure 1 depicting the isogloss for /aI/ glide-weakening were labeled as "Southern." Forty-two participants, 39 of which were recruited from North Carolina State University, fit this description. These participants were mostly from Raleigh or other areas within North Carolina, though a few were from other areas within the South. Participants who grew up outside of this area were labeled as "Non-Southern." Forty-two participant also fell into this category, all of whom were recruited from the University of Pennsylvania. Notably, the baseline vowel productions of Southern and Non-Southern participants are not much different. This is somewhat expected since Southern participants were recruited from a large state university and a majority grew up in Raleigh or other urban centers in the South, in which the Southern Vowel Shift has been reversing (Dodsworth and Kohn, 2012). Participants were all native speakers of American English and reported no speech or hearing impairments. Five participants had to be excluded for various reasons including not completing all portions of the experiment or technical issues with the participant's recording. Data is therefore analyzed for 40 Southern participants and 39 Non-Southern participants. All participants were given either course credit or a $10 Amazon gift card in exchange for their participation.

A linear mixed effects regression model was fit to the dataset of elicited /aɪ/ tokens using the lmerTest (Kuznetsova et al., 2017) package in R R Core Team (2015), which provides p-values for lmer model fits (lme4, Bates et al., 2015), using Satterthwaite's degrees of freedom method. Normalized F1 at the glide is the dependent variable. Fixed predictors were included based on likelihood ratio test significance ($p < .05$). The final model is shown in Table 1. Fixed effects included in the final model are as follows:

- **Phase:** Categorical predictor referring to experiment phases baseline (reference), exposure, and post-exposure, treatment coded

- **Voice:** Categorical predictor referring to the between-subject conditions that differ in the

Random effects:	Groups	Name	Variance	Std.Dev.		
	Participant	(Intercept)	0.075	0.274		
	word	(Intercept)	0.027	0.166		
	Residual		0.309	0.556		
Fixed effects:	Estimate	Std. Error	df	t-value	P-value	
(Intercept)	0.036	0.059	111.74	0.611	0.542	
Phase(Exposure)	0.079	0.028	4479.48	2.841	0.004	**
Phase(Post-Exposure)	0.016	0.028	4468.77	0.558	0.577	
Voice(Midland)	0.058	0.068	96.92	0.851	0.397	
Dialect(Non-Southern)	-0.161	0.064	75.74	-2.522	0.014	*
Duration	-1.838	0.174	1312.36	-10.562	<2e-16	***
Frequency	0.069	0.034	55.64	2.045	0.046	*
Phase(Exposure):Voice(Mid)	-0.096	0.04	4468.65	-2.376	0.018	*
Phase(Post):Voice(Mid)	-0.02	0.04	4466.12	-0.493	0.622	

Table 3.1: Linear mixed effects regression model predicting normalized F1 at the glide (80%) for the Pilot Experiment.

model talker voice, treatment-coded. Levels include Southern (reference) and Midland.

- **Dialect:** Categorical predictor referring to participant dialect background, treatment coded with levels Southern (reference) and Non-Southern
- **Duration:** Continuous predictor referring to token duration, z-scored
- **Frequency:** Continuous predictor referring to lexical frequency, z-scored
- **Phase*Voice:** Two-way interaction between Phase and Voice. Since these terms are included in an interaction and are treatment coded, the main effect of Phase considers the Southern condition only, and the main effect of voice considers just the baseline phase data.

By-participant and by-item random intercepts were also included in the model.

As expected, the model reports significant main effects of duration and frequency, such that more frequent words are produced with weakened glides (Est. = 0.069, p = 0.004) and longer vowels tend to have stronger glides (Est. -1.838, $p < .001$). The main effect of Phase shows that participants produce higher F1 (a lower, more Southern-like glide) after Exposure to the Southern model talker (Est. = 0.079, p = 0.004),[2] but the baseline and post-exposure phases do not differ significantly, suggesting participants return to their baseline after Exposure. There is also a signifi-

[2]Note that the main effect of Phase only refers to the Southern Condition since both predictors are treatment coded and put into the model as an interaction.

cant interaction between Phase and Voice, suggesting that these shifts from baseline to exposure are greater in the Southern Condition than in the Midland (control) Condition (Est. = -0.096, p = 0.018). Figure 3.4 illustrates these results. In the Midland Condition, participants shift very little across the three experimental phases, whereas in the Southern Condition, participants shift to produce higher F1 (lower in the vowel space) glides in the exposure phase, then return to their baselines post-exposure. As Figure 3.5 shows, this effect is primarily driven by Southern participants. Of those in the Southern Voice Condition, Non-Southerners shift little in the expected direction, with a slightly lower /aɪ/ glide in the exposure phase, but Southerners exhibit larger shifts. However, a three-way interaction between Phase, Condition, and Dialect does not significantly improve the model.

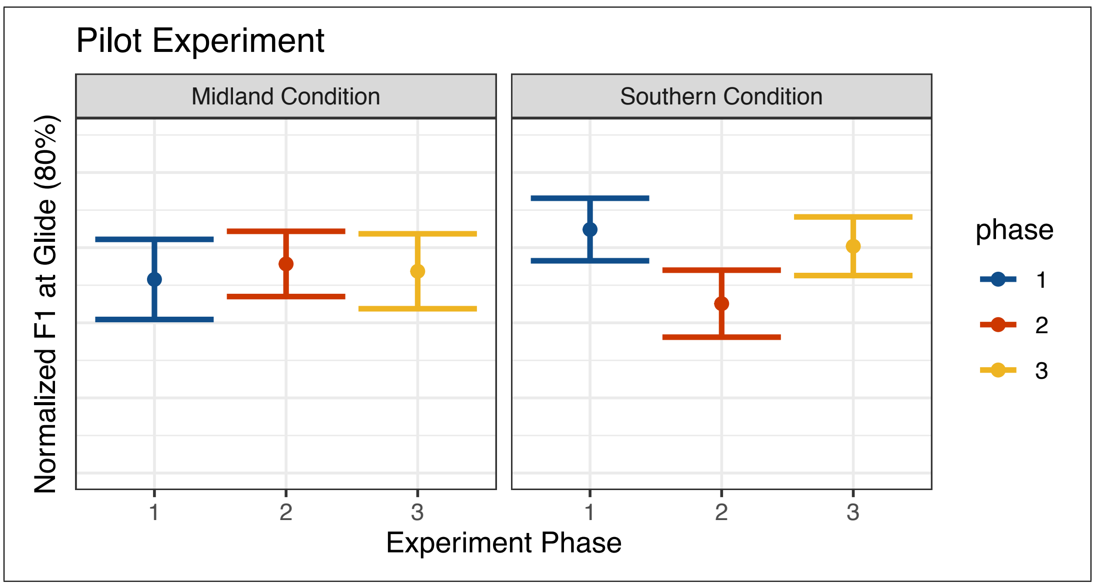

Figure 3.4: Shift across phases by Voice Condition. Mean values with 95% confidence intervals for all tokens.

The pilot also examined multiple potential acoustic measures for determining whether convergence has occurred and results suggested that height of the glide (F1 at 80%) is the best measure. Euclidean distance between the nucleus and glide and a combined measure of vowel height/frontness (normalized F2-F1) were some of the other measures considered. Since shifts in F2 were generally minimal across experimental phases, and F2 has been found to show greater variability to begin with, particularly in regard to coarticulatory effects (Fowler, 2005; Fowler and Brancazio, 2000), F1 appeared to be a better and more precise measure of convergence shifts. Shifts in F1 at the nucleus were observed, such that participants produced lower, more centralized F1 throughout the

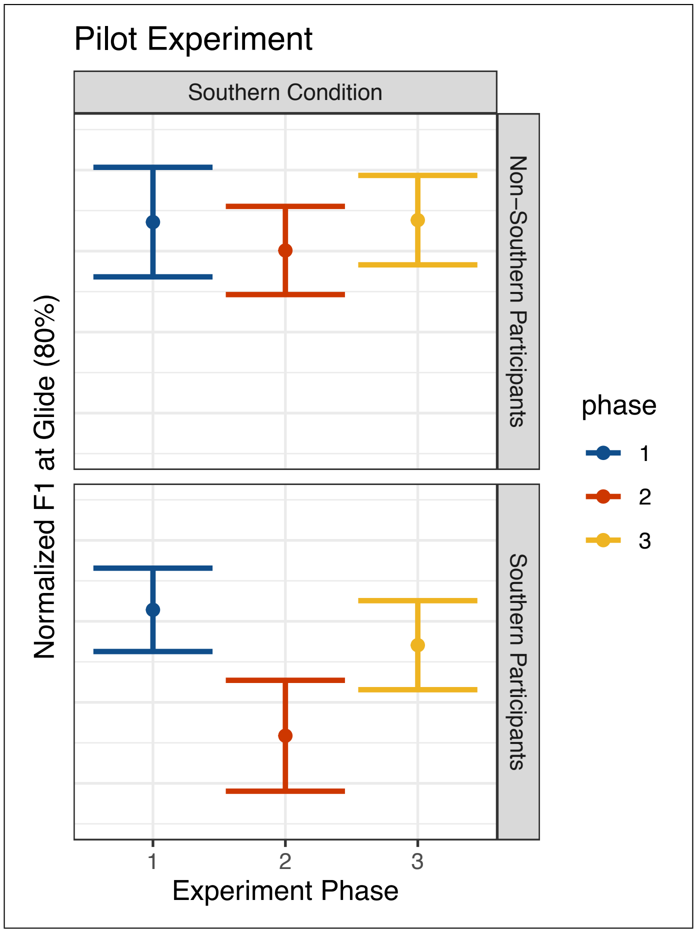

Figure 3.5: Shift across phases for only the Southern Voice condition, by participant Dialect. Mean values with 95% confidence intervals for all tokens.

course of the experiment, regardless of the Condition, suggesting participants generally produced vowels with more reduced nuclei as the experiment progressed, likely as a result of fatigue. For this reason, Euclidean distance which takes nucleus measurements into account was not used. Height of the glide was ultimately chosen as the best measure, as it was most stable across the course of the control condition and didn't appear to show general effects of undershoot as the experiment progressed. This measure seems to best allow for isolating effects of the experimental manipulation without introducing variability stemming from global shifts in production across the experiment, regardless of Condition. Notably, Experiment 1 was revised to use female model talkers (with the Southern talker being from Mississippi), and a different subject pool of participants, most of whom come from the online Prolific platform. Showing that results replicate with a new talker and subject pool, and when administered online rather than in the lab, is important to establish since subsequent experiments build off of this same paradigm. The results from Experiment 1, based on this pilot, are reported below.

3.4 Experiment 1 Results

3.4.1 Expectation-driven convergence to Southern glide-weakened /aɪ/

A linear mixed effects regression model was fit to the data using the lmerTest (Kuznetsova et al., 2017) package in R (R Core Team, 2015), which provides p-values for lmer model fits using Satterthwaite's degrees of freedom method. Normalized F1 at the glide (80%) was the dependent variable. Fixed and random effects were included based on likelihood ratio test significance. Only effects that significantly improved the model ($p < .05$) are included. The final model is shown in Table 3.2. Fixed predictors are as follows:

- **Phase:** Categorical predictor referring to the experimental phase, a within-subjects manipulation. Levels include Baseline (reference level), Exposure, and Post, treatment coded.
- **Voice:** Categorical predictor referring to the model talker dialect, a between-subjects condition with levels Southern (reference level) and Midland, treatment coded.
- **Dialect:** Categorical predictor referring to a participant's dialect background, with levels Southern (reference) and non-Southern, treatment coded.
- **Duration:** Continuous predictor referring to token duration, z-scored.
- **Frequency:** Continuous predictor referring to lexical frequency, z-scored.
- **Age:** Continuous predictor referring to participant age, z-scored.
- **Conscientiousness:** Continuous predictor referring to participants' score on the Conscientiousness subsection of the Big Five Inventory, z-scored.
- **Phase*Voice:** Two-way interaction between Phase and Voice. Because Phase is treatment coded with Baseline as the reference level, and Voice is treatment coded with Southern as the reference level, the main effect of Phase is restricted only to the Southern condition, and the main effect of Voice is restricted only to the Baseline Phase.

Scaled residuals:						
	Min	1Q	Median	3Q	Max	
	-4.376	-0.612	-0.025	0.591	5.632	
Random effects:						
Groups	Name	Variance	Std.Dev.	Corr		
Participant	(Intercept)	0.082	0.286			
	Phase(base)	0.049	0.221	-0.12		
	Phase(post)	0.055	0.235	-0.21	-0.05	
	duration	0.064	0.254	0.14	0.26	0.1
Word	(Intercept)	0.174	0.417			
	Dialect(South)	0.002	0.049	-0.35		
	duration	0.008	0.087	-0.7	0.92	
Fixed effects:						
	Estimate	Std. Error	df	t-value	p-value	
(Intercept)	0.157	0.073	184.285	2.148	0.033	*
Phase(Exposure)	0.116	0.042	101.385	2.740	0.007	**
Phase(Post)	0.040	0.054	102.447	0.742	0.460	
Conscientiousness	-0.018	0.036	101.356	-0.492	0.624	
Voice(Midland)	-0.079	0.072	99.818	-1.102	0.273	
Duration	-0.445	0.033	188.036	-13.657	<2e-16	***
Dialect(Non-Southern)	-0.229	0.059	99.452	-3.907	0.0002	***
Frequency	0.116	0.042	79.659	2.764	0.007	**
Age	0.084	0.029	96.900	2.926	0.004	**
Phase(Exposure):Conscientious	0.088	0.031	100.465	2.871	0.005	**
Phase(Post):Conscientious	0.084	0.039	102.491	2.153	0.034	*
Phase(Exposure):Voice(Mid)	-0.150	0.061	98.921	-2.473	0.015	*
Phase(Post):Voice(Mid)	-0.020	0.078	102.106	-0.260	0.795	

Table 3.2: Linear mixed effects regression model predicting normalized F1 at the glide (80%) for Experiment 1.

- **Phase*Conscientiousness:** Two-way interaction between Phase and Conscientiousness.

The model also includes by-Participant random slopes for Phase and duration and by-Item random intercepts for Dialect and duration.

As predicted, and consistent with the results of the pilot experiment, there are significant main effects of Frequency and Duration, such that shorter tokens and higher frequency items are produced with higher F1, indicating a weaker (more Southern-shifted) /aɪ/ glide. The model shows a significant main effect of Phase (and because the two-way interaction of Phase and Voice is included in the model, this main effect only considers the reference level Southern Voice Condition.) Participants shift from their baselines to produce more Southern-like glides during the Exposure Phase (Est = 0.116, p = 0.007). However, this shift does not last into the Post-Exposure phase. The Post-Exposure phase in the Southern Voice Condition does not differ significantly from the baseline

(Est = 0.040, p = 0.460), suggesting that participants return roughly back to their baseline productions after exposure. Figure 3.6 plots the mean values of Normalized F1 at the glide (80%), with 95% confidence intervals for all tokens. While those in the Midland Voice Condition exhibit no significant shifts throughout the course of the experiment, those in the Southern Voice Condition shift from their baselines to produce a lower /aɪ/ glide in the Exposure phase, then shift back up toward their baselines again in the Post-Exposure phase. Further, the interaction between Phase and Voice is significant in the model (Est = -0.150, p = 0.015), confirming that the difference from Baseline to Exposure phase in the Southern Voice Condition is significantly greater than in the Midland Voice (control) Condition. Also note that there is no significant main effect of Voice, confirming that, as expected, participants in the two randomly assigned Voice Conditions do not differ in their baseline productions of /aɪ/.

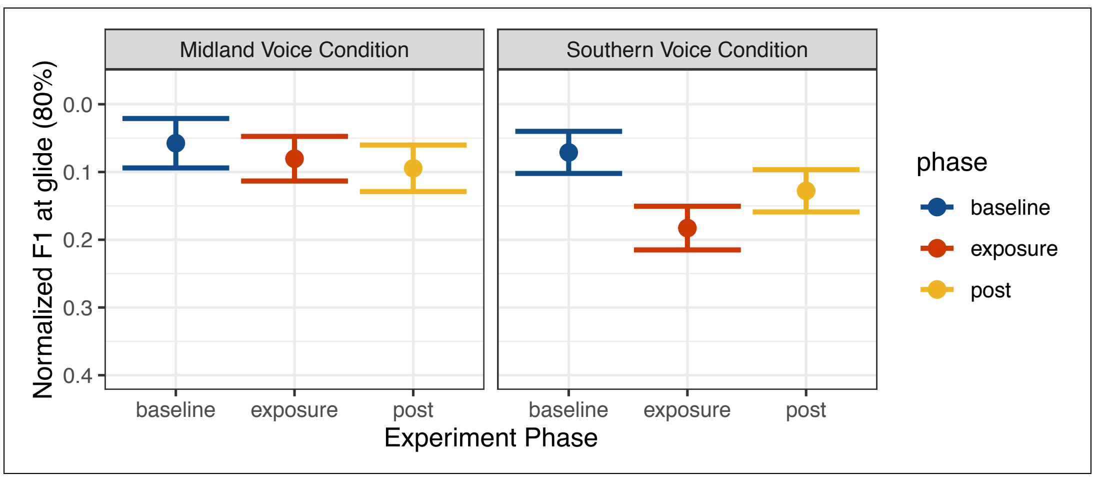

Figure 3.6: Shift across phases by Voice Condition. Mean values with 95% confidence intervals for all tokens.

The model also shows a significant main effect of Dialect, such that Non-Southerners produce lower F1 generally (Est = -0.229, p = 0.0002), which is the expected dialect difference between the two groups. This main effect collapses across the three experimental phases, and may pick up not just on baseline differences between the two groups but also differences in the degree of shift toward Southern-like productions in the exposure phase. Figure 3.7 breaks down the aggregate effect shown in Figure 3.6 by participant dialect background. While both groups exhibit the same general trend, with a lower glide in the Exposure phase then a shift back towards the Baseline in the Post-Exposure phase, these effect appears to be driven primarily by Southerners, and Non-Southern participants appear to exhibit these shifts to a lesser degree. All three phases show overlapping confidence intervals and mean shifts are in general smaller for Non-Southerners. However, neither a two-way interaction between Phase and Dialect nor a three-way interaction between Phase, Voice, and Dialect, significantly improves the model and are therefore not included.

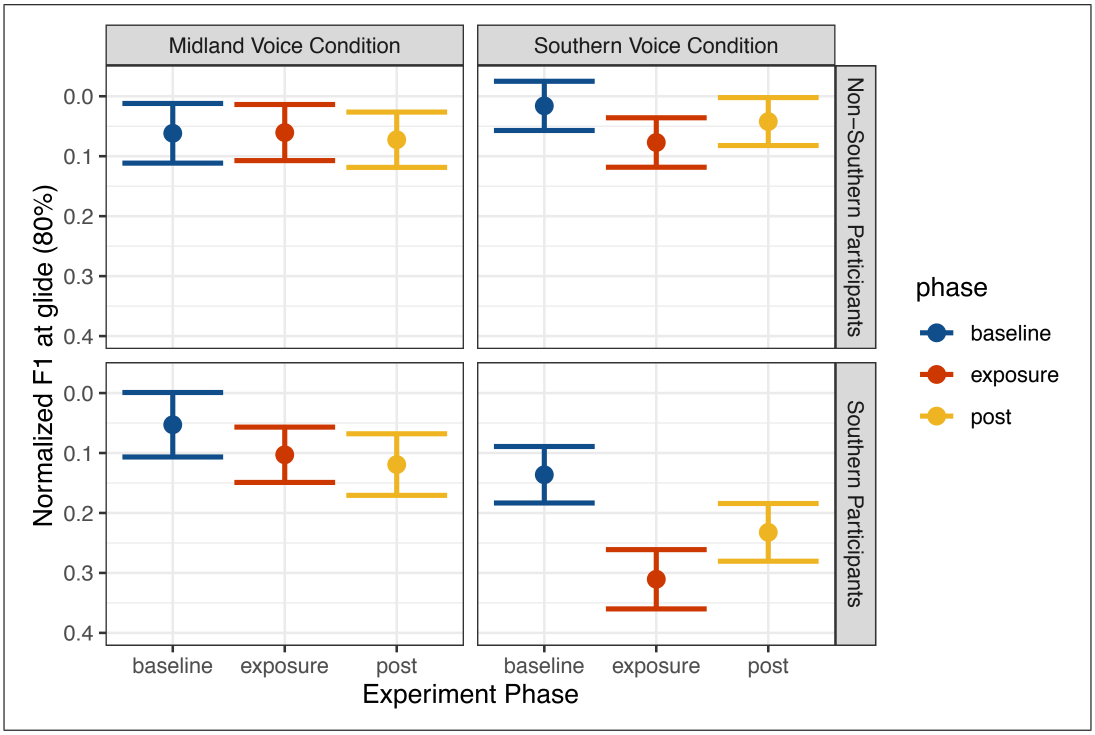

Figure 3.7: Shift across phases by Voice Condition and participant Dialect. Mean values with 95% confidence intervals for all tokens.

All demographic variables (Age, Gender, Education, Race) were also tested in the model, both as main effects and in a two-way interaction with Phase and three-way interaction with Phase and Voice. The only demographic factor that significantly improved the model was a main effect of Age. However, there is no significant interaction between Age and Dialect or Age and Phase, suggesting that older speakers produce higher F1 (a more glide-weakened vowel) across the board, but that this is not a result of older speakers being more likely to be Southern-shifted or a reflection of older speakers being more willing to converge. Note that there are no significant effects of the other demographic variables, nor any significant interactions between these variables and Phase or Phase*Voice.

We may wonder at this point whether Non-Southerners shift somewhat less than Southerners because they are shifting in other dimensions not captured in the height of the glide (F1 at 80%

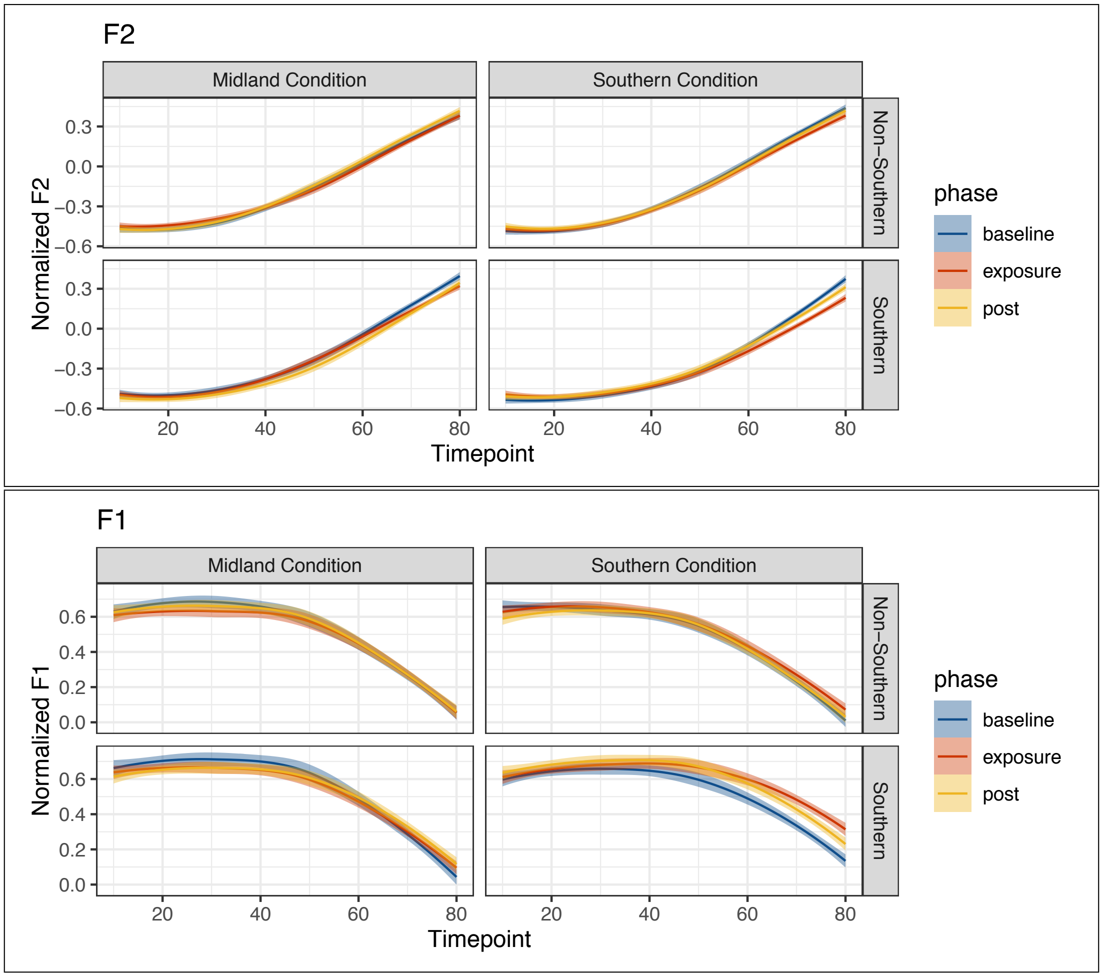

Figure 3.8: Formant trajectories for F1 (bottom) and F2 (top).

measure). Since Non-Southerners may be less familiar with the Southern dialect, attempts at converging toward glide-weakened /aɪ/ may be less accurate and not target the height of the glide. Formant trajectories for both F1 and F2 are visualized in Figure 3.8. Vowel measurements were taken at the 10%, 20%, 30%, 50%, 70%, 80%, and 90% points in the vowel, though 90% values are excluded here due to low token count from missing values or inaccurate values that were excluded as outliers. The graph plots loess curves across these 6 time points for all tokens. As we can see, there are only minimal shifts in F2 at the glide across phases, with Southerners shifting toward marginally lower F2 values throughout the course of the experiment, consistent with a more Southern production. However, similar (smaller) shifts in this same direction occur for Southerners in the Midland condition, suggesting this may be partially due to global shifts in speech as the experiment progresses. Non-Southerners exhibit no noticeable shifts in F2 at any point in the course of the vowel. In F1, we see that Southerners in the Southern Condition produce somewhat different values only in the latter half of the vowel across the three phases of the experiment. F1 values toward the end of the vowel are somewhat higher (consistent with a more Southern production) in the exposure phase than in the baseline phase, and post-exposure values lie intermediate between the two, consistent with what we see in Figure 3.7. These differences across phases are greatest at the 80% mark, suggesting our measure of glide height used throughout the analysis best captures any convergence shifts present.

3.4.2 Individual differences in accommodative behavior

Of course, not all participants adhere to the patterns exhibited in the aggregate and converge toward the Southern model talker. In the Southern Voice Condition, Southern participants exhibit more uniform convergence than Non-Southern participants. As Figure 3.9 shows, roughly half of non-Southerners shift in the opposite direction (i.e., they diverge from the Southern model talker) but only 8 (or 30%) of Southern participants diverge. While Southerners do not appear to make *greater* shifts toward the model talker (mean shift of convergers is +.26 for Southerners, compared to +.25 for Non-Southerners), *more* participants appear to be exhibiting similarly sized shifts, suggesting that individual differences may play a role in whether a participant converges. The remainder of this section investigates various individual traits that may facilitate individual convergence to the Southern model talker.

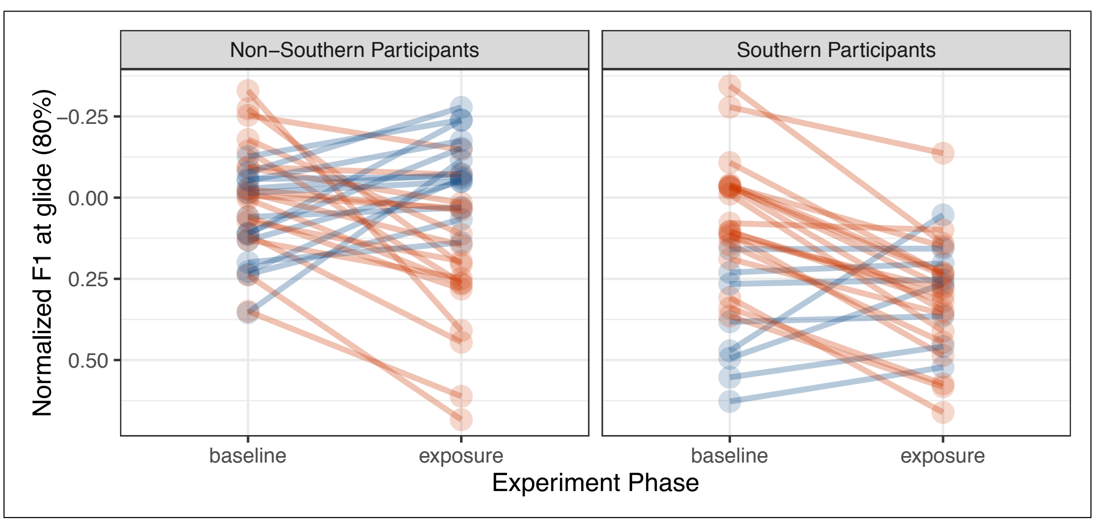

Figure 3.9: Individual Differences. Red indicates convergence. Blue indicates divergence.

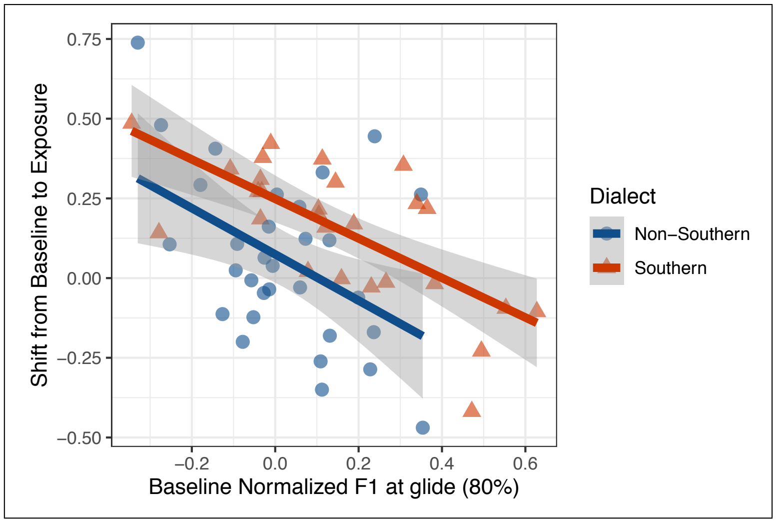

Figure 3.10: Convergence shift as a function of participants' baseline productions. Those with more Southern vowels to begin with shift less.

Dialect Distance: Of the 8 Southerners who fail to converge toward the Southern model talker, these individuals tend to be more Southern-shifted to begin with; in fact, the five participants with the most Southern baseline production of /aɪ/ all fail to shift toward the model talker. Baseline production therefore seems to play a role in the extent to which an individual converges. Figure 3.10 plots each participant's average shift in /aɪ/ glide height from the baseline to the exposure phase, as a function of their baseline /aɪ/ glide height. Shift is calculated for each participant as the mean of all /aɪ/ tokens in the exposure phase minus the mean of all /aɪ/ tokens in the baseline. Positive values indicate a shift *toward* the Southern model talker, while negative values indicate divergence, and values at 0 indicate no shift. Those with higher, less Southern glides to begin with shift more than those who already produce somewhat weakened glides, regardless of dialect background, though this effect is greater for Southerners ($R = 0.697$, $p = 0.00005$) than for Non-Southerners ($R = 0.464$, $p = 0.009$). We might conclude then that a convergence target must be different enough from a participant's own production in order for convergence to occur. This may be because of floor effects (i.e., it is impossible to imitate something you are already doing), or it may have to do with greater attention paid to novel and salient features. Note that even Non-Southerners who do not produce traditional Southern monophthongal /aɪ/ still converge less when this target is closer to their own production, suggesting that a convergence target does not only have to be non-identical in order for convergence to occur, but that it must be sufficiently distant from an individual's own production.

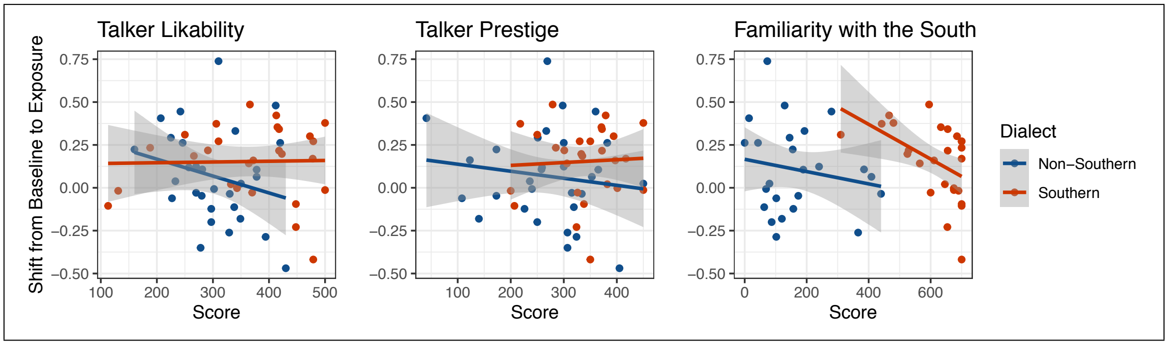

Figure 3.11: Correlation of attitudinal measures and convergence, by dialect background. Graphs show values only for those in the Southern Voice Condition.

Familiarity with the South: Figure 3.7 showed that the aggregate pattern of convergence is stronger for Southern participants than for Non-Southerners, though this Dialect difference does not reach statistical significance in our model. Still, a gradient measure of familiarity with the South may reveal a correlation between exposure to the Southern Dialect and a more nuanced distinction between dialect groups. A "Familiarity with the South" score was determined for each participant based on their responses to the follow-up survey at the end of Experiment 1. Participants rated their agreement with seven statements about their familiarity with the South[3] on a scale of 1 to 100 using a slider bar in Qualtrics, and their responses for all questions were summed to produce a single "Familiarity" score. The rightmost facet of Figure 3.11 plots this Familiarity score against Shift from baseline to exposure phase, by Dialect group. Contrary to our predictions, this plot reveals no correlation between familiarity with the South and degree of convergence for Non-Southerners, and a surprising inverse correlation for Southerners. However, note that those with higher familiarity with the South scores are also those who are the most Southern-shifted, and as we saw in Figure 3.10, those who are already Southern shifted tend to converge less because they are already similar to the model talker.

Attitudes toward the talker: The follow-up survey also gauged perceived qualities of the model talker, which can be grouped into "Likability" qualities and "Prestige" qualities. Participants rated these statements in the same way as the Familiarity with the South statements described above.

[3]The seven statements participants judged were as follows: *I have lived in the South, I am familiar with Southern speech, I have friends who are from the South, I speak like I am from the South, My parents have a Southern accent, My colleagues have a Southern accent*, and *I can imitate a Southern accent well.*

Likability qualities include *happy, kind, friendly, relatable*, and *someone I would be friends with*, and Prestige qualities include *intelligent, wealthy, professional, attractive,* and *educated.* Scores for each category were summed, producing a single "Likability Score" and "Prestige Score" for each participant. The left two facets of Figure 3.11 reveal that, surprisingly, attitudes toward the talker did little to mediate participants' convergence rates. In fact, Talker Likability actually negatively correlates with degree of shift for Non-Southern talkers (though this trend is not statistically significant).

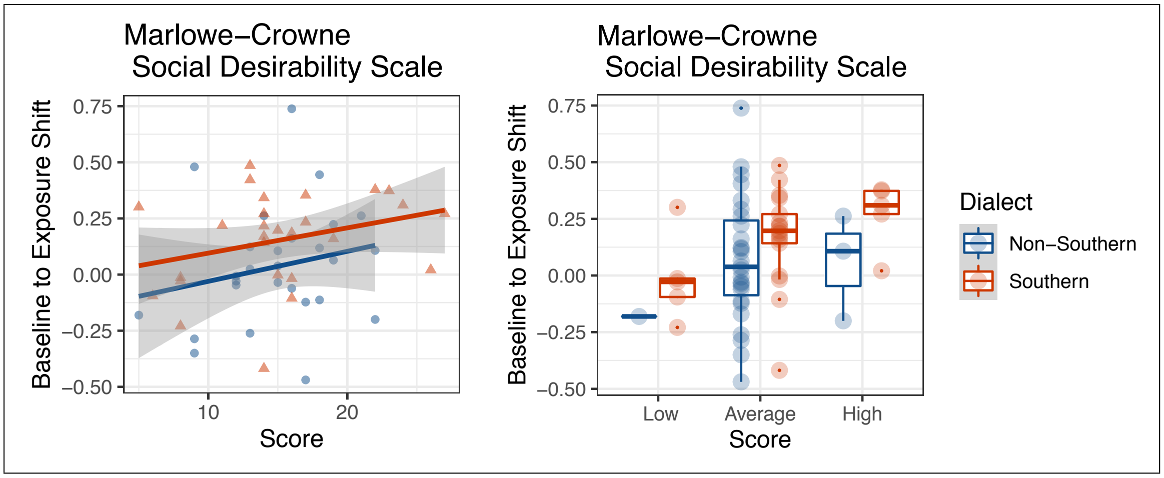

Figure 3.12: Marlowe-Crowne Social Desirability Scale scores, by dialect background. Graphs show values only for those in the Southern Voice Condition.

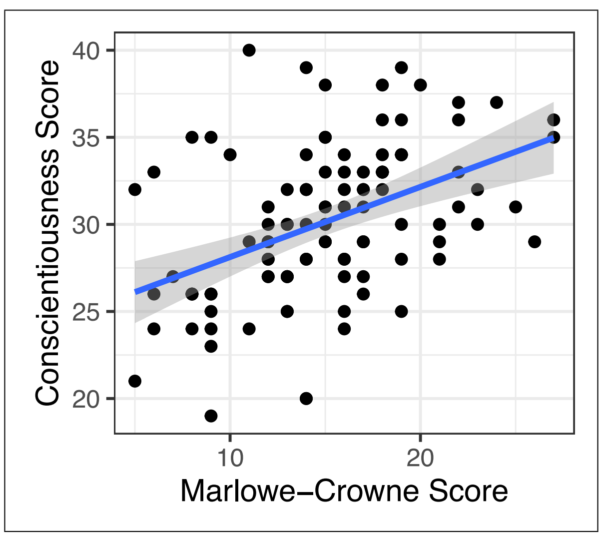

Figure 3.13: Correlation between participants' Marlowe-Crowne scores and Conscientiousness scores.

Cognitive-style/Personality: While attitudes toward the particular model talker do not seem to

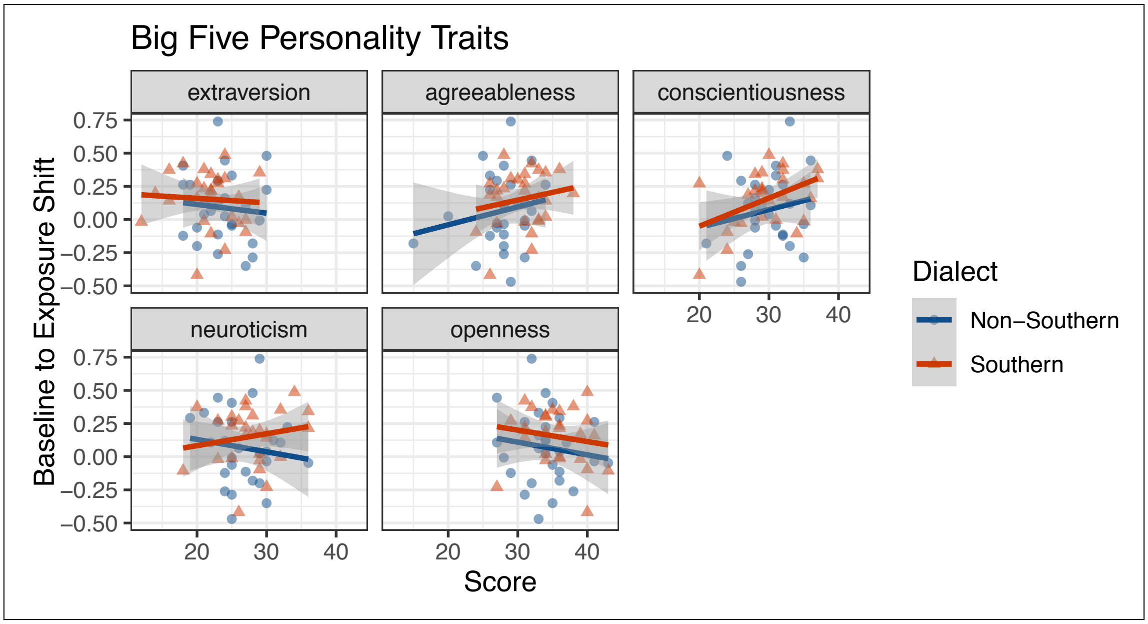

Figure 3.14: Correlation between Big Five personality dimensions and degree of convergence, by dialect background. Only participants who were in the Southern Voice Condition.

predict convergence rates, there is perhaps some evidence that desire for positive social evaluation facilitates convergence. Participants also completed the Marlowe-Crowne Social Desirability scale, which gauges a participant's tendency to respond to others in ways they deem as socially desirable. The survey consists of 33 True/False responses, which are scored on a scale of 0-33. While the majority (2/3) of participants tend to fall within the "Average" range of responses (scores 9-19), about 1/6 of survey respondents tend to score "Low" social desirability (scores 0-8), while the other 1/6 score "High" social desirability (scores 20-33) (Crowne and Marlowe, 1960). As Figure 3.12 shows, there is a moderate positive correlation between social desirability and convergence for both dialect groups. The right facet breaks these gradient scores down into the three groups used to interpret scores on the survey. The right facet of Figure 3.12 shows that 4/5 of the low-scorers diverged from the model talker, while 7/8 of the "high" scorers" converged toward the model talker. As predicted, divergers tend to be those who care less about whether their actions are perceived as socially desirable, while high convergers tend to be those who want to be viewed as socially desirable. Such a finding confirms the results of Natale (1975b) and Natale (1975a) and suggests that social motivations contribute to the tendency for an individual to converge. Note, however, that while this correlation exists, it is not included in the model. This is because scores on the social desirability scale positively correlate with another personality measure—conscientiousness on the BFI. Further, the predictors only significantly improve the model if one is added (not both), and conscientiousness appears from AIC comparison to be a better predictor of F1 height. The correlation between Social Desirability scores and Conscientiousness scores is shown in Figure 3.13 (R= .44, p = 2.272e-06). This correlation, while not predicted originally, is unsurprising. Items that are perceived as socially desirable and therefore are included on the Social Desirability Scale (e.g., "Before voting, I thoroughly investigate the qualifications of all the candidates") are very similar to the items used to gauge Conscientiousness (e.g., ("Does a thorough job"). While other personality items may not have clear positive/negative associations (e.g., "I am talkative" as a measure of extraversion), conscientiousness is clearly tied to social desirability. Likewise, those who have an increased desire to appear socially desirable would answer BFI survey questions such that they would achieve high conscientiousness scores as well (regardless of whether this reflects their actual conscientiousness). Both measures therefore seem to indicate an increased tendency for modifying behavior for social acceptance.

As Figure 3.14 shows, conscientiousness is positively correlated with degree of shift for both Southerners and non-Southerners. Indeed, the model shows a significant interaction between Phase and Conscientiousness for the Exposure (compared to baseline) phase (Est = 0.088, p = 0.005) and the Post-exposure phase (Est = 0.084, p = 0.034). This means that, those with higher conscientiousness scores shift to a greater degree in the exposure condition, and do not shift back to their baseline as much in the post-exposure phase (likely because there is further for them to shift from the exposure to post-exposure in order to reach their baseline). While a three-way interaction between Phase, Conscientiousness, and Voice did not significantly improve the model, note that the correlation between Conscientiousness and degree of shift occurs for the Southern Voice Condition (R = 0.3, p = 0.024), but not for the Midland Voice Condition (R = 0.207, p 0.145).

Other BFI measures notably do not correlate with degree of convergence. In fact, openness, which has been shown to positively correlate with convergence, shows a slight trend in the opposite direction for both dialect groups, and neuroticism, which has also previously been shown to correlate with convergence, shows opposite trends for Southern and Non-Southern speakers, neither of which is significant. Finally, scores on the Autism spectrum quotient were also investigated as possible predictors of convergence. Specifically, the Attention Switching subscore has been shown to positively correlate with convergence (Yu et al., 2013). However, no such correlation was found in the present study.

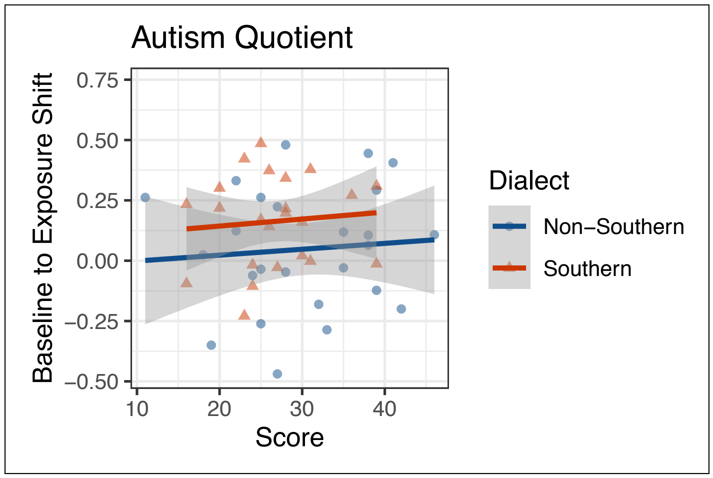

Figure 3.15: Correlation between aggregate AQ and degree of convergence shift, only for those in Southern Voice Condition.

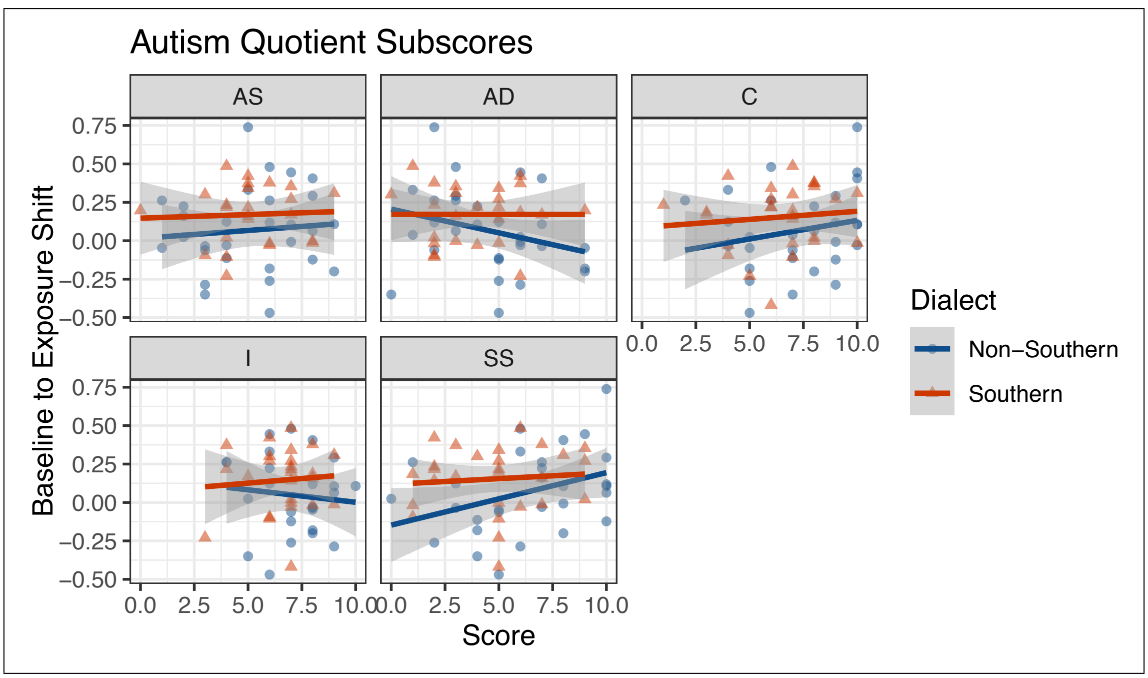

Figure 3.16: Correlation between AQ subscores and degree of convergence shift, only for those in Southern Voice Condition.

3.4.3 Shifts toward observed vowels

Because tokens of other vowels were elicited from participants, primarily for normalization purposes, this also allows for a post-hoc analysis of convergence toward other vowels, though bear in mind that the Ns for vowels other than /aɪ/ are quite low and this study was not designed to test convergence to other vowels specifically. We can ask whether we find any tentative evidence for convergence to the entire system of Southern-shifted vowels, including vowels observed from the model talker within the context of the experiment, or whether speakers shift only toward glide-weakened /aɪ/—arguably the most salient feature of Southern U.S. English.

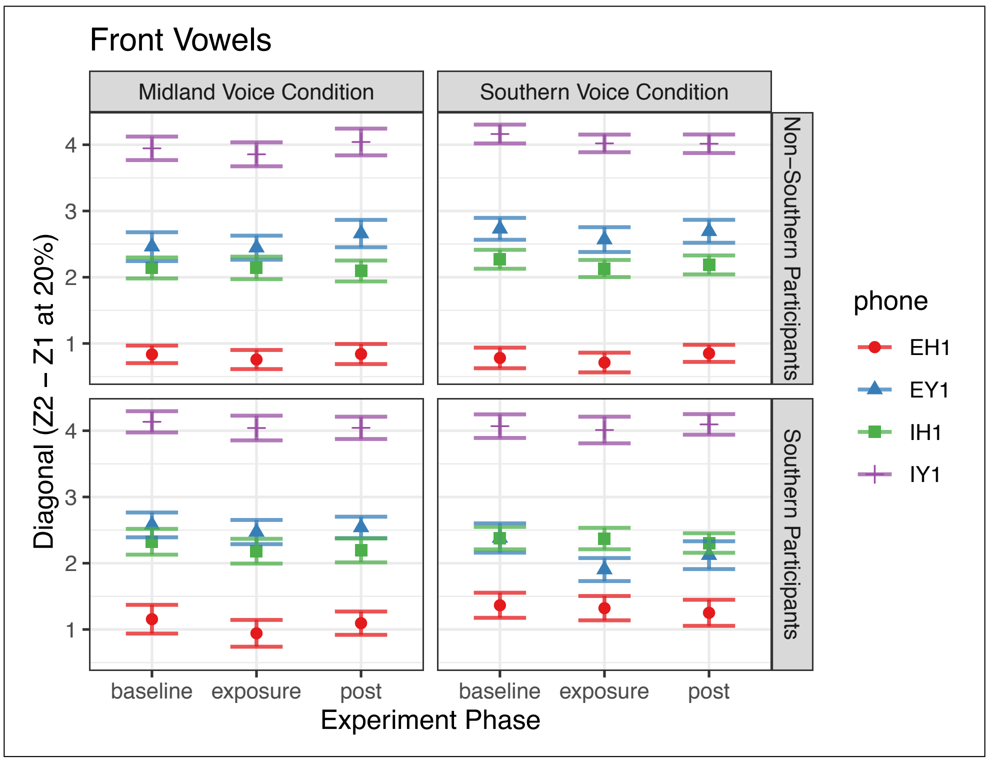

Figure 3.17: Front vowels: mean values for all tokens with 95% confidence intervals.

Figure 3.17 plots mean diagonal values with 95% confidence intervals for all tokens of just the front vowels. The diagonal (Z2-Z1) is often used to simultaneously measure the frontness and height of the vowels along the front diagonal of the vowel space. The higher on the plot, therefore, the higher and fronter in the vowel space. Front vowels appear relatively stable across conditions—there is little difference between the minimal shifts in the Midland condition and those in the Southern condition. However, there is one noticeable exception. The /e/ vowel lowers and

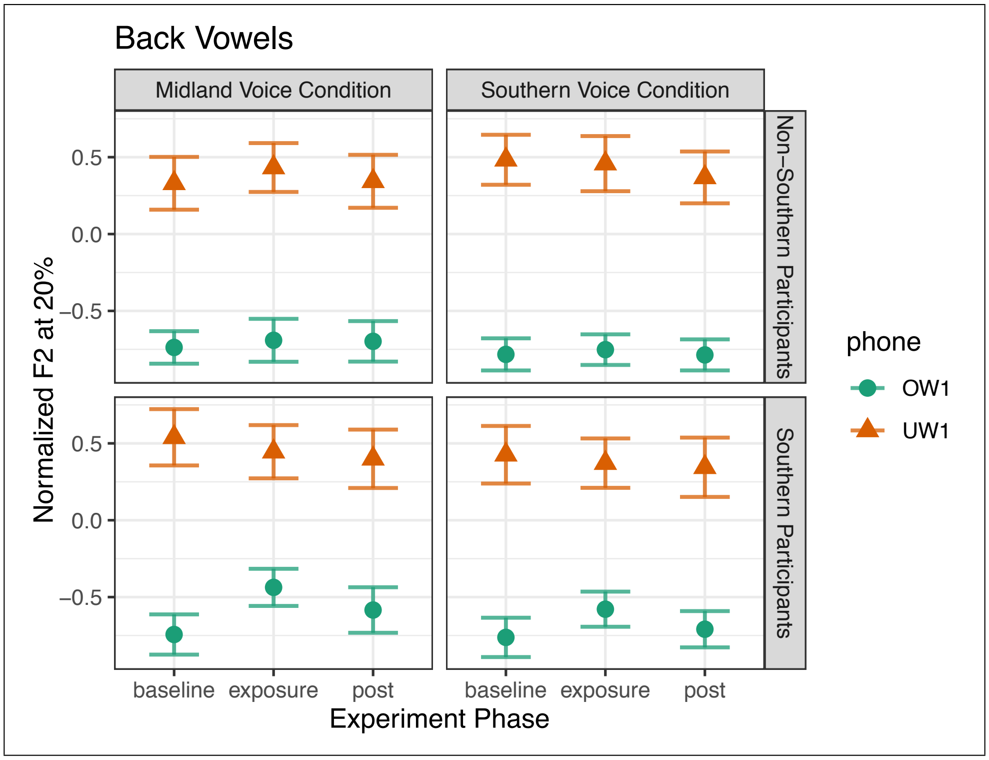

Figure 3.18: Lack of shift for back vowels: mean values for all tokens with 95% confidence intervals.

backs along the front diagonal in the exposure condition, then shifts back up in the post-exposure condition, but only for Southerners in the Southern condition. Notably, Non-Southerners do not appear to be shifting to any of the vowels they actually observed from the Southern model talker throughout the course of the experiment. For the back vowels shown in Figure 3.18, there is also a lack of shifting toward Southern variants. Southern participants produce slightly more fronted /o/ vowels in the exposure phase, but this holds for both the Midland and Southern Voice Conditions and does not appear to be a result of the experimental manipulation. None of the shifts for vowels other than /aɪ/ are statistically significant.

3.5 Discussion

This experiment used a novel Word Naming Game paradigm intended to elicit convergence from participants who interact with—but crucially do not hear target items from—a model talker. Using this design, Experiment 1 has provided evidence that speakers do exhibit expectation-driven convergence. Participants in the aggregate produced more Southern, glide-weakened tokens of the

/aɪ/ vowel after hearing a Southern talker who never produced any instances of /aɪ/ throughout the course of the experiment. Further, shifts toward glide-weakened /aɪ/ were significantly greater for those who heard a Southern talker than for those in the Midland talker condition, who did not exhibit such shifts; this suggests that shifts toward glide-weakened /aɪ/ were truly a result of the experimental manipulation and cannot be attributed to general shifts in production (e.g., due to fatigue) as the experiment progressed. Further, the results of Experiment 1 mirror those of the pilot experiment; showing that these findings replicate with a different model talker and subject pool suggests that expectation-driven convergence is a real and reliable linguistic behavior. Findings of expectation-driven convergence also suggest that the Word Naming Game paradigm is suitable for eliciting convergence experimentally and may prove a useful method for tackling other questions across linguistics and related disciplines.

The existence of expectation-driven convergence is of interest for several reasons, particularly because it indicates a range of other behaviors that we do not directly observe. (1) For one, initial associations must be formed via exposure. Individuals, at various points in their lives, must have been exposed to /aɪ/-glide weakening from a Southern talker, and eventually would have formed mental associations between this linguistic variant and the group of people who use it. The amount of exposure required to be able to form such associations is likely somewhat high; in order to associate /aɪ/-glide weakening with a particular type of talker, multiple exposures to this variant from multiple talkers likely have occurred, and in order for these associations to be retained in memory and not decay over time, activation of these associations would likely need to occur periodically. In other words, in order to simply form the initial association between /aɪ/-glide weakening and a particular type of talker, a substantial amount exposure is likely necessary, which may vary widely based on factors such as the dialect region in which an individual has lived. (2) When exposed to the Southern model talker, individuals must be able to successfully *recognize* based on the model talker's speech that this a Southern talker (or that this talker uses other Southern-shifted variants) and (3) this recognition must ultimately lead to activation of the associated glide-weakened /aɪ/ variant, based on the associations formed in step 1. (4) Finally, an individual must incorporate this activated linguistic variant (or a shift *toward* this variant) in their own speech, which entails remembering and targeting detailed phonetic information, either by automatic or more agentive socially-motivated mechanisms. When we observe expectation-driven convergence then, this entails a whole host of other behaviors occurring in the past and present that we do not directly observe. Importantly, these behaviors have implications for theories of the mechanisms behind convergence and the relationship between perception and production. For one, models that rely on automatic perception-production feedback loops or automatic imitation of motor gestures in which production is derived directly and automatically from perceptual input are not straightforwardly able to explain expectation-driven convergence, which demonstrates that individuals are able to converge toward a linguistic form that is not directly observed in the input. This is not to say that *no* instances of convergence are accomplished by these mechanisms, or that the mechanisms behind expectation-driven convergence cannot involve automatic perception-production links. Rather, this simply means that convergence can target non-locally observed forms. The existence of expectation-driven convergence is still compatible with an automatic perception-production relationship, where crucially the triggering item that is perceived is not the same as the production target, though these two component are somehow associatively linked. Alternatively, expectation-driven convergence may be a result of more agentive, socially-motivated processes. Finally, the existence of expectation-driven convergence suggests that detailed phonetic forms be stored in memory in order to be targeted in later speech production, and that these phonetic forms are in some way linked to social knowledge.

In addition to establishing expectation-driven convergence as an empirically observable phenomenon and determining the suitability of the novel Word Naming Game paradigm for eliciting linguistic convergence, the results of the experiment also allow us to conclude several things about the nature of expectation-driven convergence. The first is that it is relatively short-lived. Participants shift back to their baselines almost immediately after the exposure phase, and there is no lingering evidence of convergence post-exposure. There could be several reasons for this ranging from lack of motivation to converge to lack of a necessary convergence trigger in the Post-Exposure phase. That is, participants could discontinue convergence in the Post-Exposure phase because they have returned to reading clues rather than hearing them, and with their pseudo-interlocutor no longer present, participants may have simply lacked the social motivation to converge. Alternatively, while hearing a Southerner triggered convergence in the Exposure phase, in the absence of this trigger, participants may have no longer been primed to converge.

Experiment 1 also demonstrated that participants exhibited considerable individual differences in production behavior, and not all participants in the Southern Voice Condition converged as expected. One potential source of these differences comes from experience with the Southern dialect—only 30% of Southern participants failed to converge to the Southern talker, while nearly half of non-Southern participants failed to converge. This is not surprising, since only those familiar with Southern speech should be able to imitate it based on expectations alone. It is likely the case that it is much more important for the forms a speaker converges toward to be familar to them when there are no instances of these forms available as models within the context of the experiment. Bell (1984) also picks up on this potential complication, noting that, "for outgroup design, however, absence of feedback has crucial consequences for a speaker's performance. The speaker has no access to the outgroup, and therefore lacks adequate models of outgroup speech...they have to overcome ignorance both of a target speech community to which they may have no access, and of a target variety which they may have never heard spoken natively" (p. 190). While dialect background may offer promising directions for understanding whether a given individual will converge to a non-standard dialect feature, we see no significant interaction between participant dialect and experimental phase in the statistical modeling, and participants' self-reported familiarity with the South scores do not reliably predict degree of shift toward the Southern talker. In fact, Southerners who self-reported the greatest familiarity with the South actually converged the least, though this is likely because these speakers were also the most Southern-shifted to begin with and did not need to shift much—or at all—in order to match the speech of the Southern model talker. While more Southern participants converged to the Southern talker, it is also the case that the least Southern-sounding participants in either dialect group tended to converge the most. That is, participants whose baseline productions of /aɪ/ were already somewhat weakened did not converge as much as those with very diphthongal /aɪ/ vowels. While this may at first seem contradictory to the finding that more Southerners converged, it actually fits nicely with previous findings that the most phonetically distinct tokens (those which are novel and therefore most salient) elicit greater convergence, but only if the target is within a speaker's repertoire (i.e., if speakers have experience producing such forms) (e.g., Walker and Campbell-Kibler, 2015). We also find some support for Bell's (1984) claim that variation exhibited through convergence to an interlocutor of another group does not not exceed the variation exhibited between these groups. Indeed, the difference in /aɪ/ glide production between Midlanders and Southerners (Est = -0.229, p = 0.0002) is greater than convergence shifts (Est = 0.116, p = 0.007).

One somewhat surprising observation is that attitudes toward the talker and attitudes toward the south did not seem to predict degree of convergence for either Dialect group, despite a range of previous studies suggesting that attitudes do play a role in convergence. There are a number of potential reasons for this. For one, the attitudinal measures used here were limited to assessing talker likability, talker prestige, and likability of the South. It is possible that the attitudinal measures that may have mattered most for convergence were not assessed in the follow-up survey.

There is, however, some evidence that cognitive-style/personality traits may play a role in convergence behaviors. Low-scorers on the Marlowe-Crowne social desirability scale tend to diverge from the Southern model talker, while High-scorers tend to converge. Unsurprisingly, those with little desire for social approval do not converge as much as those with a heightened desire for social approval. If convergence is largely an automatic process, it may be the case that low-social desirability individuals in fact try to maintain their own linguistic norms and actively inhibit any automatic convergence effects. Alternatively, if convergence stems from more agentive social motivations, it may be that moderate/high social desirability indicates the motivation required to trigger convergence. Further, social desirability as measured by the Marlowe-Crowne Social Desirability scale correlated positively with individuals' Conscientiousness scores on the BFI. Conscientiousness ended up being a slightly better predictor of convergence, likely for the same reasons of social approval. Findings that a desire for social approval positively correlates with convergence mirror findings of Natale (1975a) and Natale (1975b), who found a positive correlation between convergence and Marlowe-Crowne Social Desirability Scale scores, as well as Lewandowski and Jilka (2019), who found a positive correlation between convergence and Neuroticism scores, which they interpreted as indicating a higher need for social approval. The present experiment, however, finds no correlation between neuroticism and convergence. Further, the results of the present study notably contrast with previous findings that Openness scores on the BFI and Attention Switching scores on the AQ positively correlate with convergence because both signal greater focus and attention to phonetic detail. This lack of correlation is perhaps unsurprising since there is no phonetic signal for participants to attune to in the course of the experiment. Therefore, greater attention to phonetic detail (at least locally) would not be predicted to increase convergence toward non-observed glide-weakened /aɪ/. Whether Openness and Attention Switching play a different role in input-driven convergence will be examined in the next experiment.

Finally, I investigated whether convergence occurred to any of the other observed Southern-shifted vowels in the experiment. While Southerners shifted minimally to produce a more Southern-like /e/ vowel in the Southern voice condition, this effect was not significant and no other notable shifts were observed. There could be a number of reasons for the lack of convergence to observed vowels, ranging from the salience of the variant to the lack of power since token Ns of non-/aɪ/ vowels were significantly lower. Convergence toward expected vs. observed variants will be probed more directly in the next experiment, which contains both an /aɪ/–present and /aɪ/–absent condition.

Chapter 4

Experiment 2: The Perceptual Underpinnings of Expectation-Driven and Input-Driven Convergence

This chapter reports the results of Experiment 2, which investigates the perceptual underpinnings of convergence and compares expectation-driven convergence and input-driven convergence directly. I use an individual differences design to compare each participant's behavior across a range of tasks, allowing for the perception-production relationship to be probed at the individual level. In addition to testing the relationship between perception and production, Experiment 2 is to designed to contribute to our understanding of how different individual domain-general propensities may facilitate convergence. By comparing individual difference predictors of input-driven and expectation-driven shifts in both perception and production, we can begin to understand how these two behaviors are related. For instance, if Social Desirability scores positively correlate with expectation-driven convergence, as observed in Experiment 1, but not input-driven convergence, we might conclude that the two types of convergence are driven by somewhat different mechanisms, where social motivations are more relevant for motivating expectation-driven convergence. If individual measures tend to correlate with expectation-driven and input-driven convergence in the same direction, we might conclude that these types of convergence are driven by the same mechanisms.

Of particular interest in investigating the perceptual underpinnings of shifts in production is the role of perception in *divergence*; while the predicted perceptual underpinnings of *convergence* are clear—if there is a relationship between perception and production, shifts in perception toward a variant would induce corresponding shifts in production—the predicted role of perception in *divergence* is not as clear cut. We can imagine several possible ways in which perception shifts and convergence/divergence may be related, which would yield different patterns in the combined perception-production data: (1) While convergence may result from perceptual shifts *toward* a variant, divergence might stem from perceptual shifts *away* from a variant. The idea of perceptual divergence has been proposed, for instance, by Walker et al. (2018). If Experiment 2 finds evidence supporting this relationship, this would indicate maximally linked perception and production behaviors, where production would fall out automatically from perception with no room for inhibition or facilitation of automatic shifts via external factors. It's unclear, however, what would motivate perceptual divergence, particularly if perceptual responses are an automatic consequence of exposure. (2) Another possible relationship we may observe is that divergence in production may stem from the *same* perceptual shifts as convergence. That is, perceptual shifts *toward* a recently heard variant may be an automatic consequence of exposure, while shifts in production may be more controlled and subject to external influence; participants who have perceptually integrated a novel variant may then *converge* toward this variant, but they could also *diverge* or fail to shift at all. In other words, while production shifts may *require* perceptual shifts toward that variant, a perceptual shift may not be sufficient for eliciting convergence. This would suggest a greater role for factors such as attitudes and social motivations in facilitating or inhibiting convergence. (3) Finally, we might also observe no obvious relationship between perception and production. A lack of correlation may indicate several things, including that (a) shifts in perception and production are variably influenced by external factors, such as social motivations, that may obscure any relationship, (b) the tasks used to elicit perception and production are not measuring what they intended to, or (c) perceptual shifts and production shifts result from largely distinct mechanisms. The investigation of individual perception-production links in Experiment 2 will therefore have implications not only for our understanding of the relationship between perception and production, but also the nature of linguistic divergence, and its relationship to convergence.

To investigate shifts in production, I use a modified version of the same Word Naming Game task presented in Experiment 1, which tests whether individuals converge by producing a variant they might reasonably expect from a Southern talker—glide-weakened /aɪ/—when they never actually observe the /aɪ/ vowel from the model talker vs. when they do. As in Experiment 1, the task includes a Southern voice condition, in which the Southern talker never uses the /aɪ/ vowel (referred to here as the Southern /aɪ/–absent Condition), as well as a Midland (control) condition. In order to directly compare expectation-driven and input-driven convergence, a Southern /aɪ/–present condition was also added, in which the same Southern talker *does* produce glide-weakened /aɪ/ vowels throughout the course of the experiment.

Perceptual shifts toward glide-weakened /aɪ/ were investigated using an auditory lexical decision task, in which participants judged sound clips spoken by a Southern-shifted talker as real "words" or "non-words" after hearing one of the three types of exposure described above (Southern /aɪ/–absent, Southern /aɪ/–present, or Midland (control)). Target items on the lexical decision task include /aɪ/-words produced with glide-weakened /aɪ/, such that a word like "bribe" may sound like a non-word "brahb". I predict higher "word" response rates for these target items when participants have recently been exposed to a Southern talker—and even higher response rates if the Southern talker produces glide-weakened /aɪ/ in the exposure phase.

Specific research questions are as follows:

1. Do participants shift their perceptual boundary to encompass more glide-weakened tokens of the /aɪ/ vowel after hearing a Southerner who does or does not produce the /aɪ/ vowel?
2. Are perceptual shifts mediated by a participant's dialect background?
3. Do perceptual behaviors predict convergence toward (or divergence from) the same variant?
4. Do expectation-driven convergence (/aɪ/–absent condition) and input-driven convergence (/aɪ/–present condition) differ in degree of convergence, longevity of convergence, likelihood of convergence across dialect groups, or individual differences predictors?

4.1 Background

This section reviews the literature informing the research questions and predictions of Experiment 2. Subsection 4.1.1 reviews findings of perceptual learning of both observed and expected linguistic behavior, highlighting an auditory lexical decision paradigm for investigating perceptual learning that will be used in Experiment 2. Subsection 4.1.2 then explores the literature on the relationship between production and perception, particularly that of relevance to linguistic convergence and reports on previous findings exploring the relationship between perception and production.

4.1.1 Perceptual Learning

In Experiment 2, in order to investigate the perceptual underpinnings of convergence, I use a perceptual learning paradigm to test how recent exposure to a novel dialect or variant results in perceptual category adjustment. In linguistics, perceptual learning refers to the process whereby perceptual categories are adjusted after some sort of exposure to a novel pronunciation. For instance, a listener may perceptually adjust the category boundaries between two phonemes after being exposed to a speaker who produces a given category in a non-familiar way. This is also sometimes also referred to as adaptation or recalibration.

In experimental investigations of perceptual learning, category adjustment is often induced lexically; for example, listeners may be influenced to hear an ambiguous segment as /f/ if it is presented in a context in which /f/ is the only segment that produces a real word, such as *gira?* for *giraffe.* This is what Norris et al. (2003) did in their seminal study on lexically-guided perceptual learning. One group of listeners heard an ambiguous sound between [s] and [f] in place of /s/, while another group heard the sound in place of /f/, in the context of a lexical decision task. Listeners who heard the ambiguous sound in place of /s/ showed an /f/-/s/ category boundary closer to /f/ (e.g., categorizing more tokens as /s/) than those who heard the ambiguous sound in place of /f/, suggesting that listeners expanded their category boundaries to encompass the ambiguous sound they had heard. This type of lexically-induced perceptual learning has been replicated in a number of studies (e.g., Eisner and McQueen, 2005; Reinisch and Holt, 2014; McQueen et al., 2006; Eisner and McQueen, 2006; Kraljic and Samuel, 2007). In addition to continuum-categorization tasks, another recently used measure of perceptual learning is ease of lexical access, measured by likelihood or speed to accept words with newly learned pronunciations as 'words' (Maye et al., 2008; Weatherholtz, 2015). For instance, Maye et al. (2008) found that inducing learning by passively exposing participants to a passage of *The Wizard of Oz,* in which all of the front vowels had been lowered, resulted in increased identification of novel words (not heard in the passage) with synthetically-lowered front vowels as "words" in a lexical decision task. This method of measuring perceptual learning has also been used to investigate perceptual responses to expected forms. Under a similar paradigm, Weatherholtz (2015) found that exposing participants to a novel chain shift resulted in greater "word" responses for items with shifted vowels in a lexical decision task. Further, listeners exposed to only a subset of a previously unheard vowel chain shift were able to fill in gaps and generalize to phonemes that were not a part of training, suggesting listeners had learned a pattern of co-variation among vowel categories.

4.1.2 Perception-production relationships

Cognitive mechanisms accounting for linguistic convergence often assume tight perception-production linkages, such that convergence is an automatic consequence of perception. The episodic theory of speech perception and production, for instance, suggests that production targets are derived directly from linguistic perception. Episodic lexical entries are stored in memory with detailed phonetic information attached to each. "Echoes" or memory traces of previously heard lexical items are activated during production of the same lexical item, and production targets are derived from the distribution of these activated traces. Convergence occurs when a person hears a particular word form and the distribution from which they draw a production target has shifted shifted to encompass recently heard forms. Accounts of convergence appealing to structural priming mechanisms also assume tight perception-production relationships, and are considered a route for automatic perception influences on production. The Interactive Alignment Model (Pickering and Garrod, 2004), for instance, suggests that individuals align their linguistic representations at all levels of linguistic structure due to automatic priming mechanisms that increase intelligibility for both speakers. Under a priming account, a recently heard form later "primes" a similar or related form, facilitating its access and promoting its usage. Other accounts under which perception and production are tightly linked include gestural-based accounts of speech perception such as Motor Theory, which predict convergence on the basis of gestural perception. In Motor Theory (Liberman and Mattingly, 1985), speech perception units consist of articulatory gestures rather than acoustic properties. In other words, gestures are recovered during speech comprehension. Similarly, Direct Realist Theory (Fowler, 1986) considers gestures to be the units of speech perception, but unlike Motor Theory argues that perception occurs in terms of actual (rather than intended) vocal tract gestures, and may thus be in a better position to explain linguistic convergence (Galantucci et al., 2006). Recently, this link between perception in terms of gestures and convergence has been made more explicit by models proposing that language users internally simulate the motor commands for speech they perceive by running forward models of their own speech production (Pickering and Garrod, 2013; Gambi and Pickering, 2013). Support for gestural accounts of convergence come from evidence that phonetic imitation occurs when model utterances are presented only visually (Gentilucci and Bernardis, 2007; Miller et al., 2010). Despite theories linking perception and production tightly, and a wide range of supporting experimental evidence, these relationships are not always borne out in studies comparing experimentally induced changes in perception to changes in production. For instance, Mitterer and Ernestus (2008) provide evidence against both phonetically detailed perception-production links and gestural perception. They found no latency in perceiving two different variants of Dutch /r/. That is, participants with a natural aveolar /r/ who responded to a uvular /r/ (or vice versa) showed no latency in identification. They suggest that the link between perception and production is minimal, but is phonological and abstract rather than phonetically detailed, and also note that this explains why production is not an automatic consequence of perception.

While much work has been done on the "perception" component of perceptual learning, little work has been done on whether or how perceptual learning adjustments in particular go on to influence production, and the work that has been done does not produce consistent results. For instance, Kraljic et al. (2008) found that, even though individuals exhibited large perceptual learning shifts in the /s/-/ʃ/ category boundary after exposure to a novel pronunciation, these individuals did not necessarily exhibit equivalent shifts in production. They concluded that "While such perceptual changes might prove to be necessary for production changes, they do not seem to be sufficient" (p. 15). Alternatively, Lehet and Holt (2017) *did* find a stronger relationship between perceptual learning shifts and subsequent shifts in production. Participants exposed to an artificial English dialect with noncanonical use of f0 for the stop voicing contrast decreased their reliance of f0 in perception, and subsequently decreased their own use of f0 in production. As of now, all we can conclude from the range of studies examining perception-production relationships is that, while some link apparently exists, perception and production do not always align precisely. As Schertz and Clare (2019) note, "The dynamic nature of perception and production could make it such that there is too much variability to see a relationship even if it exists" (p. 11).

We know even less about the relationship between perception and production in regard to expectations. Some of the models able to account for input-driven convergence can also account for convergence to expectations. For instance, exemplar theoretic models assume that, during speech perception, both detailed phonetic information and contextual (e.g., social) information are stored in memory. Thus, when a social concept is activated, linguistic exemplars associated with the social concept may also be activated. Further, shifts in perception based on expectations have recently been observed. As mentioned above, using a similar lexical decision design as the present experiment, Weatherholtz (2015) found that listeners exposed to some vowels undergoing a previously unheard chain shift were more likely to recognize words containing shifted vowels as "words" than those without previous exposure. Further, this perceptual shift extended to previously unheard vowels, suggesting that listeners were able to extrapolate the unheard parts of the chain shift based on their partial exposure. Perceptual linguistic shifts have also been shown to be induced by non-linguistic information. Hay et al. (2006a) found that participants who all heard the same New Zealand talker judged the talker's /ɪ/ vowels as fronter (i.e., more Australian-like) when the word "Australia" was printed at the top of their answer sheet—regardless of being told they were listening to a New Zealander. Similarly, Hay and Drager (2010) found that, participants being in a room with stuffed toys representing different dialect groups (a kangaroo or koala for Australia or a kiwi for New Zealand) resulted in similar shifts in vowel categorization. Both studies have considered these findings support for exemplar-theoretic models of speech, in which social concepts activate associated linguistic forms that then influence production. Even divergence in speech perception has been reported. Walker et al. (2018) found that perceptual shifts *toward* or *away* from an Australian production of the KIT vowel were induced by reading good or bad facts about Australia. While the perception-production relationship between input-driven and expectation-driven convergence is unclear, one way to conceptualize expectation-driven convergence is as essentially input-driven, but with a longer latency between perception and production. Still, perception and production to the same variants—particularly when these variants are only *expected* but not observed—remains under-explored and is the focus of Experiment 2.

4.2 Methods

Expectation-driven and input-driven perceptual shifts are investigated using a between-subjects exposure-test auditory lexical decision design following (Weatherholtz, 2015), and corresponding convergence shifts were investigated using the same Word Naming Game design as in Experiment 1. Participants completed both the production and perception components so that behavior in both dimensions can be compared at the individual level. The details of the experimental methods are described below.

4.2.1 Participants

180 participants were originally recruited to participant in Experiment 2. However, due to a glitch that resulted in failure to save a portion of the audio recordings, additional participants were recruited, and more participants have perception data than production data. Overall, 212 participants ended up being recruited to participate, with audio data available for 178. Of the 212 participants, 161 were recruited through Prolific and completed the experiment online for $10 USD. Fifty-one participants were recruited through the University of Pennsylvania subject pool and were given course credit for their participation. Of these, 36 participated in the Language Variation and Cognition Lab in the University of Pennsylvania's Linguistics Department, and 15 participated online. Ten participants were excluded from just the production analysis for missing more than 20 clues or having poor sound quality that made formant measurements unreliable. 168 participants' production data are analyzed here. Some participants were also excluded from the perception analysis; Nine participants who had less than 75% accuracy on non-target items (i.e., fillers and non-words) were omitted. After these exclusions, combined perception-production data were available for 159 participants. As in Experiment 1, half of participants spent the majority of their school-aged years (5-18) in the U.S. South and were thus labeled as "Southern" (N=107 for perception, N=91 for production), while the remaining half spent this time in the U.S. outside of the Southern isogloss shown in Figure 2.4, and were labeled as "Non-Southern" (N=96 for perception, N=77 for production). These participants' baseline vowel measurements are shown in Figure 4.1. Though most vowel categories are produced the same across dialect groups, Southerners produce largely overlapped /e/ and /ɛ/ categories, which are generally distinct for Non-Southerners. Further, while the nucleus of /aɪ/ is

the same for both dialect groups, the glide is notably lower and somewhat backer for Southeners. This is a larger difference in baseline /aɪ/ productions between Southerners and Non-Southerners than observed in Experiment 1 (Figure 3.1). Figure 4.2 plots individual speakers means for /aɪ/ nucleus and glide measurements, and shows a significant portion of Southerners already produce lower and backer glides, approaching the territory of the /aɪ/ nucleus. In Experiment 1, there was a clear distinction between nucleus and glide productions for *all* participants. We can conclude that the sample of Southerners here are somewhat more Southern-shifted to begin with than those in Experiment 1. Participants were all native speakers of American English and reported no speech or hearing impairments.

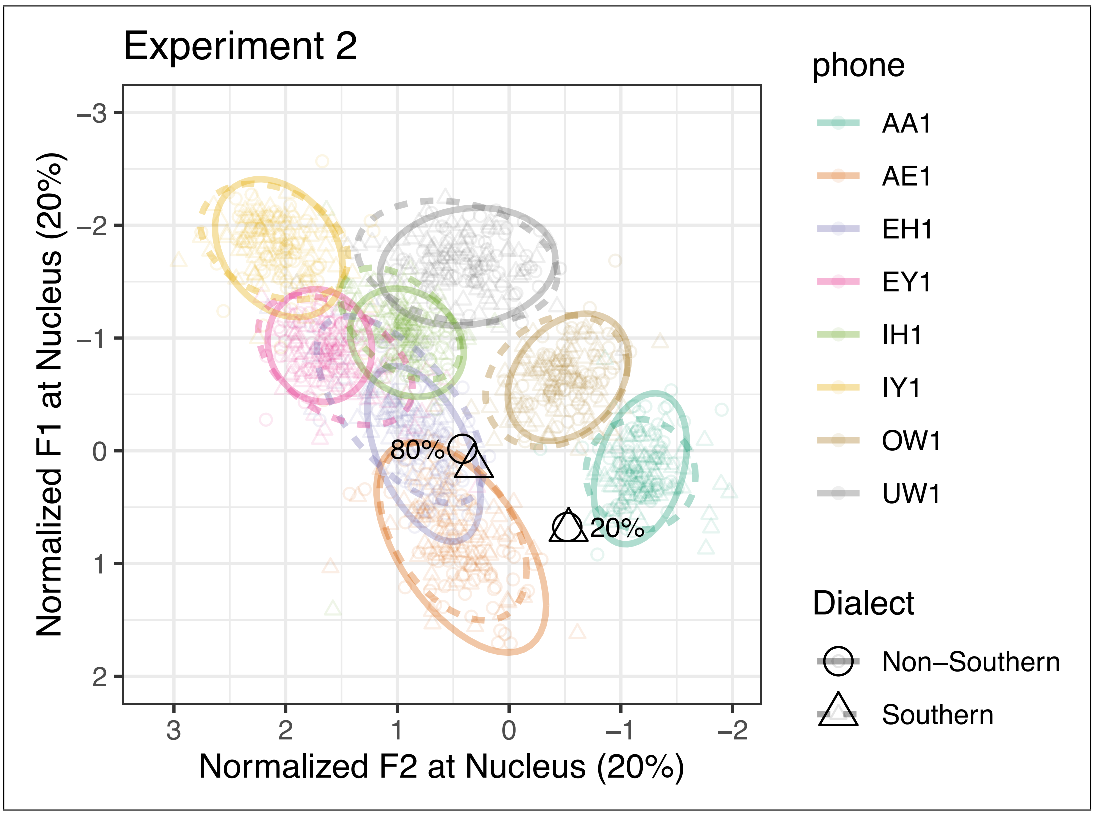

Figure 4.1: Southerner vs. Non-Southerner vowels. Each point represent participant means, and linetype indicates dialect group. Shapes indicate Southerners' (triangle) and Non-Southerners' (circle) productions of the /aɪ/ vowel at the nucleus (20%) and glide (80%).

Of the 168 participants included in the production analysis, 156 disclosed demographic data. 84 identified as female, 69 as male, and 3 as Non-binary. In regard to maximum educational attainment, 53 reported some college, 44 a Bachelor's degree, 30 a high school diploma, 17 an associates degree, 11 an advanced or graduate degree and 1 a technical certification. 122 identified as White, 13 as

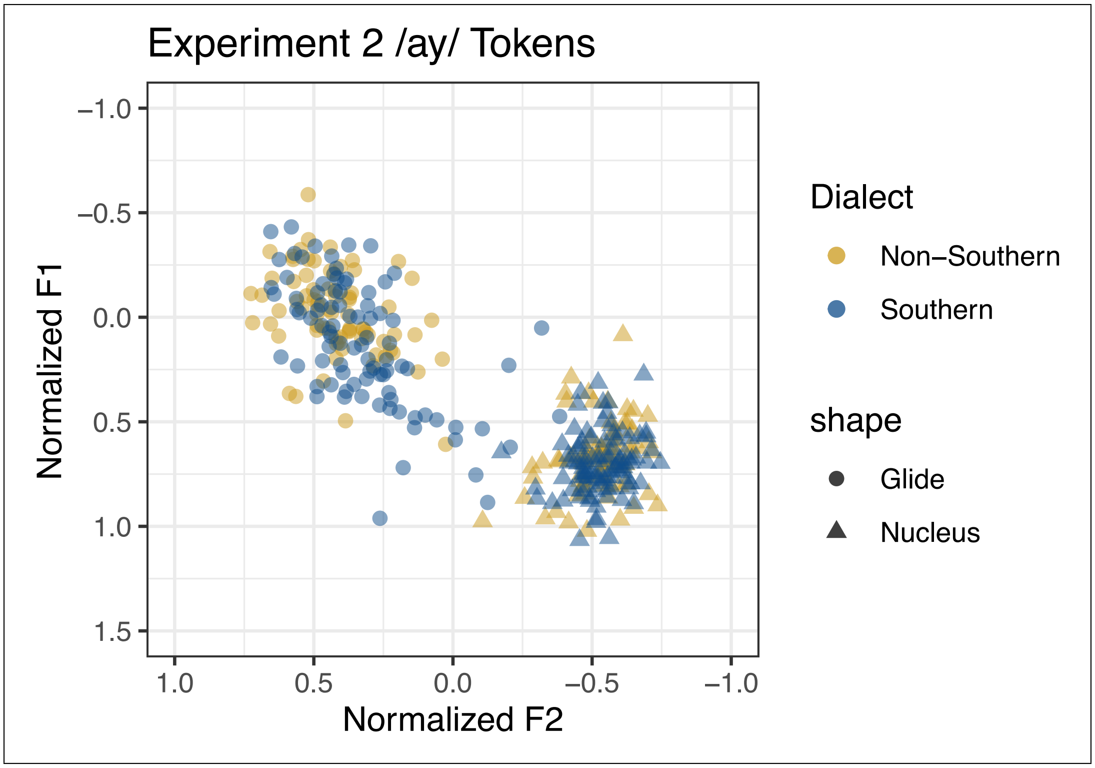

Figure 4.2: Individual speaker means for the /aɪ/ vowel, with triangles indicating the nucleus (20%) and circles indicating the glide (80%).

Black or African American, 13 as Asian or Pacific Islander, 5 as Hispanic or Latinx, and 3 as Other. Mean age is 30.66 with a range of 18–73. Of the 203 participants included in the perception analysis, 181 disclosed demographic data. 101 identified as female, 75 as male, 5 as non-binary. 66 reported some college, 46 obtained a Bachelor's degree, 35 a high school diploma, 18 an Associate's degree, 14 an advanced or graduate degree, and 2 reported "other." The racial breakdown is as follows: 135 White, 15 Black or African American, 15 Asian or Pacific Islander, 9 Hispanic or Latinx, and 5 Other. Mean age is 30.18 with a range of 18–68 years.

4.2.2 Procedure

As in Experiment 1, the experiment was designed using the PennController toolkit (Zehr and Schwarz, 2018) for Ibex and administered via participants' web browsers, and participants were recorded through their computer microphone. The initial demographic questionnaire and follow-up survey were identical to those in Experiment 1 and will not be repeated here. Further, the procedure of the Word Naming Game was generally the same. The detailed explanation of these components can be found in Section 3.2. There are a few important changes to the methods used in Experiment 1, described here.

The first change is the added perception component, which was administered immediately after the Exposure phase of the Word Naming Game and prior to the Post-Exposure phase. The Exposure phase then served the dual purpose of collecting participants' speech as part of the production measure, and serving as the exposure stimuli for the perception component. Just as in the Word Naming Game production task, the perception task uses a between-participants design with the different conditions corresponding to the stimuli heard in the Exposure phase, though I include a third condition here in Experiment 2 that was not included in Experiment 1. As in Experiment 1, participants assigned to the Midland condition heard clues from a Midland talker who never used the /aɪ/ vowel, and those assigned to the Southern (/aɪ/–absent) condition heard a Southern talker who also never used the /aɪ/ vowel. Experiment 2 adds a Southern (/aɪ/–present) condition, using the same Southern talker as the Southern /aɪ/–absent condition. In this added Southern /aɪ/–present Condition, the model talker *does* produce the /aɪ/ vowel at least once per clue, and is included in order to compare input-driven and expectation-driven convergence to the same linguistic variant directly. All clues were kept as similar as possible to those in the /aɪ/–absent conditions, with the only changes being those necessary to introduce the /aɪ/ vowel into each clue. The Southern model talker in both Southern conditions was the same Southern female talker used in Experiment 1, and the Midland talker was also the same Midland female talker as used in Experiment 1. Unlike in Experiment 1, the order of items was held constant across participants; in order to allow for individual differences comparisons, we want what participants' experience to be as similar as possible to one another.

The perception component uses an auditory lexical decision design following Maye et al. (2008); Weatherholtz (2015). In the lexical decision task, which occurred immediately after the Exposure phase, participants heard isolated words over headphones, and were instructed to decide whether what they heard was a real word or a non-word as quickly and as accurately as possible by pressing a designated key on their keyboard. Regardless of the condition, the items participants judged in the lexical decision task were all spoken by the same Southerner heard in the Southern conditions of the Exposure phase. Therefore, the lexical decision task was identical across conditions. Lexical decision items consisted of 40 target words containing /aɪ/ spoken by a Southerner who therefore produced such items with a glide-weakened vowel, such that a word such as *bribe* would sound like [bra:b]. All target items were real, monosyllabic words containing /aɪ/ in a pre-voiced or coda environment that may be confused for non-words when produced with a glide-weakened pronunciation (i.e., if /aɪ/ were substituted with /ɑ/ or /æ/). That is, a word like *bribe* could be included because *brahb* would be a non-word, but words like *ride* were not included because a glide-weakened production that sounds like *rod* or *rad* would still be a real word. Participants who reconstructed the glide-weakened pronunciation of these target items as /aɪ/ would be more likely to respond "word" while those who do not (i.e., they think they hear "brahb") would be more likely to choose "non-word." The idea is that recent exposure to a Southern talker would promote higher word-endorsement rates of these target items than exposure to a Midland talker. Items also included 60 monosyllabic filler real words (e.g., "smash") and 40 monosyllabic filler non-words (e.g., "yorch"), none of which contained the /aɪ/ vowel. The lexical decision stimuli are listed in Appendix C. Items were presented to participants in a randomized order that was held constant across participants, in order to allow for individual differences comparisons.

The lexical decision task was followed by the post-exposure phase, which was identical to that of Experiment 1. Participants also completed the same follow-up survey as in Experiment 1, containing various individual affective, cognitive-style, and personality measures. The detailed description of this task is available in Section 3.2.

4.2.3 Analysis

The production data was analyzed in the same way as in Experiment 1. Soundfiles were transcribed, forced aligned, and blindly hand-checked and corrected, and formant tracking settings in Praat were adjusted for each speaker, and each token as necessary, with defaults of 5 number of formants and 5500 max formant height for females and 5000 for males. A Praat script was used to extract duration and F1 and F2 measurements at seven time points throughout the vowel, and formant values were Lobanov normalized in R. I focus here on normalized F1 at the glide (80%), based on previous findings that this measure captured shifts elicited by the experimental manipulated but was not heavily influenced by global speech shifts as the experiment progressed. Any outliers more than 3 standard deviations from the mean were excluded on a by-participant and by-item basis, leaving 14,494 /aɪ/ tokens for analysis (an average of 86 per participant, or 29 per participant, per phase).

In the perception component, of primary interest are the "word" and "non-word" responses on the lexical decision task. Word responses were coded as 1 and non-word responses as zero. No trials were omitted. Reaction time data was also collected, but the task wasn't designed to generate meaningful reaction time data, so this will not be analyzed here.

4.3 Results

4.3.1 Production

In the aggregate, Non-Southerners behave as they did in Experiment 1. Non-Southerners, shown in the top facets of Figure 4.3 show little to no shift across the experimental phases in the Midland Condition. In the Southern /aɪ/ Absent Condition, Non-Southerners show slight shifts of roughly the same magnitude of those observed among Non-Southerners in Experiment 1; they produce glides that are lower in the vowel space (i.e., more glide-weakened or Southern-like) after hearing a Southern talker in the exposure phase. After exposure to the model talker has discontinued in the post-exposure phase, participants shift back up to their baselines. In the new Southern /aɪ/–present condition, when participants *do* hear glide-weakened /aɪ/ from the Southern model talker, participants similarly shift during the exposure phase, and this shift persists into the post-exposure phase, where these participants produce even more reduced /aɪ/ glides. While these shifts are minimal, this is consistent with our observations of Non-Southerners in Experiment 1.

Unlike in Experiment 1, Southerners (in the bottom facets of Figure 4.3, show the opposite pattern, and produce on average *less* Southern-shifted /aɪ/ vowels after hearing the model talker in the Southern Ay-Present Condition. In fact, Southerners in the aggregate appear to *diverge* from the model talker in the exposure phase, then diverge even further in the post-exposure phase. In the Southern Ay-Absent Condition, Southerners shift very little, perhaps diverging slightly from the model talker in the exposure phase and reverting back to baseline again in the post-exposure phase. While these patterns are the exact reverse of what we see from the Non-Southerners, there is some consistency in that shifts (whether convergence or divergence) in response to Southern Ay-Present stimuli continue into the post-exposure phase, while those in the Southern Ay-Absent condition are shorter-lived.

Divergence from Southerners is unexpected and inconsistent with the results of both the Experiment 1 pilot and main study. There are several potential explanations to propose at this point.

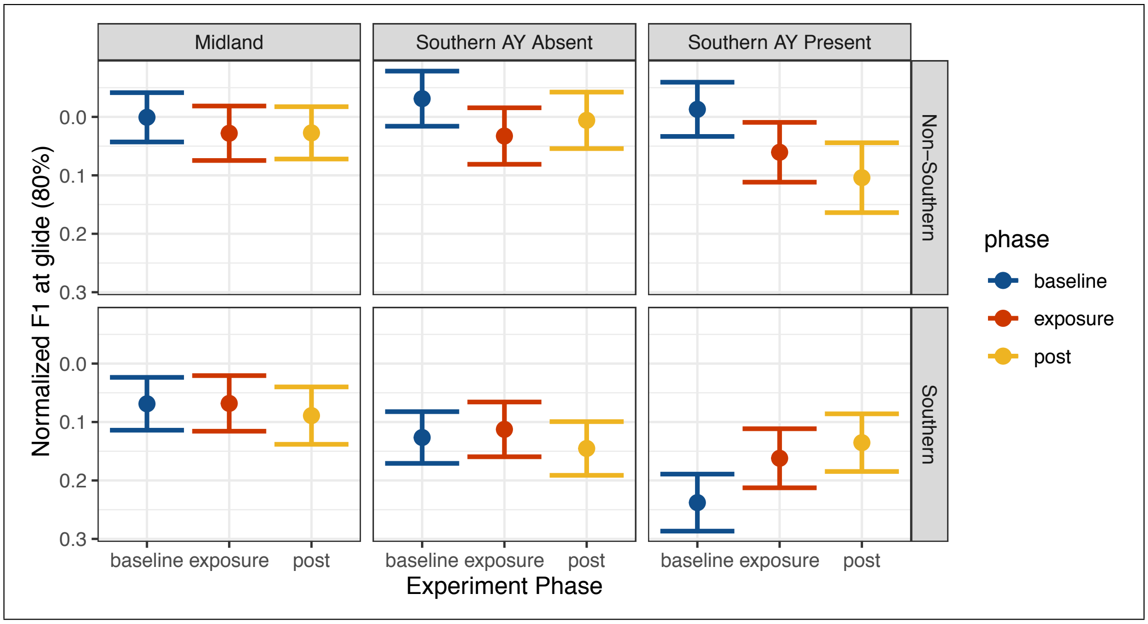

Figure 4.3: Mean normalized F1 at the glide (80%) across all tokens, broken down by participant dialect and experiment phase, with 95% confidence intervals.

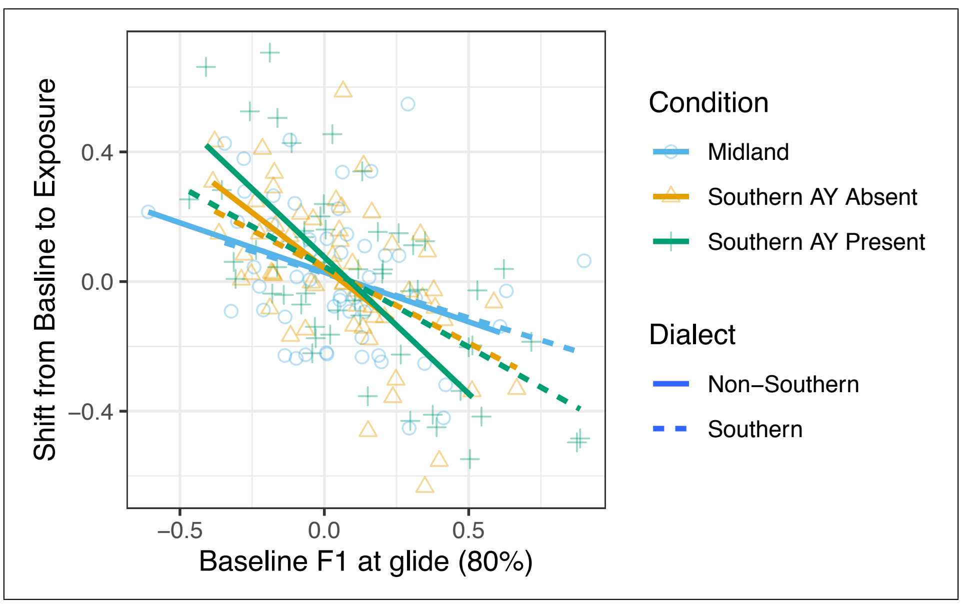

Figure 4.4: Shift from baseline to exposure, measured as the difference between the exposure and baseline phases, as a function of participant baseline production. Points represent mean values for a single participant. Line color indicates experimental condition and linetype indicates participant dialect. Shift values above 0 indicate convergence, while those below 0 indicate divergence.

The first is that the present experiment contains many more already Southern-shifted participants. Compare, for instance, the baselines of Southerners in Figure 3.1 and Figure 4.1. Also recall that Experiment 1 found a significant effect of baseline on degree of shift, showing that those with more

Southern-shifted vowels to begin with converge toward Southern speech less or actually diverge. Indeed, the correlation between participants' baseline productions and degree of shift is replicated in the present experiment as well, as shown in Figure 4.2. Participants in both Southern voice conditions shown a negative correlation between baseline and degree of shift (AyPresent: Pearson's R = -0.65 [Southern], -0.62 [Non-Southern], Bonferroni-corrected $p < .001$ [Southern] $p < .01$ [Non-Southern]; AyAbsent: Pearson's R = -0.48 [Southern], -0.67 [Non-Southern], Bonferroni-corrected $p < .05$ [Southern] $p < .001$ [Non-Southern]), such that higher (more Southern baselines) correspond to less convergence toward (or greater divergence from) the Southern model talker. While those in the Midland Condition show similar trends in the same direction, this correlation is not statistically significant for either Dialect group (Pearson's R = -0.33, Bonferroni corrected $p > .05$). A linear regression model fit to the data, described in detail below, also confirms a significant interaction between Condition and Baseline in predicting degree of Shift, such that the the effect of baseline in the expected direction is significantly greater in the Southern voice conditions than in the Midland (control) condition. The role of baseline in degree of shift therefore appears to be directly related to the experimental manipulation and is not simply a result of global speech tendencies that would apply equally across conditions.

There are several reasons why baseline productions may correlate with degree of shift in this way. For one, Southerners who already produce somewhat glide-weakened /aɪ/ may not have any room to converge (i.e., there is no "shift" necessary to produce what they are hearing from the model talker). Alternatively, participants with already Southern-shifted productions of /aɪ/ may not notice the stimuli as something available to converge toward in the first place if targets are not sufficiently different from their own speech. We may also speculate that divergence could result from attempts at phonemic contrast preservation; in other words, if individuals with already somewhat weakened /aɪ/-glides shift toward the Southern model talker, their productions may encroach upon the territory of /æ/ or /ɑ/ and threaten phonemic contrast preservation. In order to avoid contrast collapse between these categories /aɪ/, individuals may actually *hyper-correct* to preserve this contrast. Indeed, contrast preservation has been proposed as having an inhibitory effect on linguistic convergence (Nielsen, 2011). These possibilities are discussed in greater detail in the discussion section. Regardless, we see that divergence only occurs in the Southern-talker conditions and not in the Midland talking condition, suggesting that divergence shifts are in fact a response to the experimental manipulation.

There are other possibilities for why Southerners in this experiment may tend to diverge, aside from their baseline productions alone. For one, Southerners in this version of the Experiment may have had less positive feelings toward the Southern model talker, resulting in a tendency to diverge. Southerners in Experiment 2 do tend to have slightly lower Prestige ratings for the Southern talker, closer to those of Non-Southerners, while in Experiment 1, Southern participants rank the Southern talker as more Prestigious than Non-Southerners. Prestige was not, however, found to correlate with degree of shift in Experiment 1 or 2. In fact, none of the other attitudinal measures predicted convergence vs. divergence, though it is possible that Southerners in Experiment 2 happened to have differing attitudes toward the Southern talker than those in Experiment 1 in a dimension not measured in our follow up survey. A final potential contributing factor to this unexpected pattern may have to do with the fact that this version of the experiment (unlike in Experiment 1) kept item order consistent across participants in an attempt to isolate the role of individual differences in convergence behaviors. For this reason, while comparisons of relative shift *across* participants may be more reliable, absolute findings (i.e., participants tend to *converge* as a whole) may be less reliable. It is likely that multiple factors, including the three described here, are at play in contributing to the unexpected pattern we see among Southerners. Regardless, the unexpectedly high proportion of divergers results in an overall wider range of production behaviors than observed in Experiment 1, which is actually more useful in examining individual differences predictors of convergence behaviors.

Individual differences in Production

Just as in Experiment 1, there are considerable individual differences in degree of convergence (or divergence), especially since many Southerners diverge in this case, while others converge as expected. This section presents the results of the investigation into which individual differences measures may correlate with degree of convergence. This analysis is largely exploratory and has several aims. The first is to test whether the individual differences measures previously shown to predict degree of phonetic convergence are the same ones that predict convergence (or divergence) in the present study, particularly in light of recent failures to replicate past findings of individual difference correlates of convergence (e.g., Tamminga et al., 2019). We might expect some similarities, particularly in the Southern /aɪ/–present Condition that tests input-driven convergence, if such measures truly do predict convergence rates. However, we might also expect to see some differences from previous findings, particularly in the predictors relevant for the Southern Ay-Absent condition. Many of the previously discussed findings on individual predictors of convergence focus on an individual's propensity for *noticing* stimuli that trigger convergence in the context of the experiment, which would presumably be less relevant for expectation-driven convergence, which doesn't rely on noticing the target of convergence in real time. This brings us to the second main goal of this section, which is to determine whether similar individual difference measures predict convergence for both input-driven and expectation driven-convergence. This will help to tease apart whether these behaviors are governed by similar cognitive mechanisms, tap into the same domain-general capacities, and draw on similar motivations. Finally, this section has a third originally unintended goal, which is to determine whether the increased rates of divergence, particularly among Southerners, is in fact a response to the experimental manipulation, or whether this simply reflects noise in the data. If we see that Southerners and Non-Southerners' degree of shift is predicted by the same measures, but in opposite directions, this may support the idea that convergence among non-Southerners and divergence among Southerners are related responses toward the experimental stimuli, and may share some motivations or be guided by similar mechanisms.

In order to investigate the role of multiple individual differences measures in degree of convergence (or divergence) simultaneously, a linear model was fit to the data and predicts individual participant "Shift," measured as the difference between the mean normalized F1 value at the glide (80%) for all tokens in the exposure condition and those in the baseline condition. Positive values therefore indicate convergence toward the Southern talker, while negative values indicate divergence. This measure results in one observation per participant that averages across all words, so a linear model with no random effects was fit to the data. This method was chosen to avoid uninterpretable 4- or 5-way interactions and to acknowledge the fact that this experiment was designed not to determine whether convergence occurs in the aggregate (i.e., to compare values across conditions) but rather to compare *relative* shifts across participants.

Since there were 11 individual differences predictors, plus demographic factors, that may reasonably predict convergence, all of which may interact with Condition and participant Dialect, there were a large number of possible terms to consider for this model, making manual model comparison unfeasible. Further, since it was unclear which individual differences measures to expect a priori, and since this investigation is largely exploratory, a purely hypothesis-driven method of including predictors could not be utilized. For this reason, I used an automated method of model selection. I used the *regsubsets* function in the *leaps* package in R, which performs an exhaustive search for the best subsets of all possible combination of variables from a maximal model and selects one best fit model of each size. The model with the most included terms was chosen since it had the highest adjusted R-squared value. The terms provided for the subset search include all individual differences measures (5 AQ subscores, 5 BFI measures, and Marlowe-Crowne Social Desirability Score) along with the variables gender (which has been previously been proposed to influence degree of convergence) and baseline (which correlated with degree of convergence in Experiment 1), which were allowed to interact with Condition and Dialect. Rather than including the AQ total score along with the five AQ subscores, only the subscores were included since the combined subscores are equivalent to the total AQ score.

The final model chosen as the best subset of this maximal model using the *regsubsets* function includes the following predictors:

- **SocDes:** Score on the Marlowe-Crowne Social Desirability Scale, z-scored
- **AS:** Attention Switching subscore of the AQ, z-scored
- **Openness:** Openness score on the Big Five Inventory, z-scored
- **Gender:** Categorical predictor with levels male and female, sum-coded. Three participants identifying as non-binary were excluded from the gender analysis due to small Ns.
- **Baseline:** Continuous measure of participants' baseline normalized F1 at 80%
- **Voice:** Categorical predictor referring to experimental condition with levels Midland, SouthernAY, and SouthernNoAy (reference)
- **Dialect:** Categorical predictor referring to participant dialect, treatment coded with levels Southern (reference) and Non-Southern
- **Voice*Gender:** Two-way interaction between Voice and Gender

Residuals:					
	Min	1Q	Median	3Q	Max
	-1.605	-0.480	-0.021	0.453	2.008
Fixed effects:	Estimate	Std. error	t-value	p-value	
Gender	-0.218	0.111	-1.961	0.052	.
Voice(Midland):Gender(F)	0.636	0.175	3.641	0.000	***
SocDes	0.364	0.126	2.883	0.005	**
Voice(Midland):SocDes	-0.746	0.289	-2.585	0.011	*
Voice(SouthernAY):SocDes	-0.479	0.190	-2.527	0.013	*
AS	-0.457	0.170	-2.692	0.008	**
Voice(Midland):AS	0.626	0.274	2.288	0.024	*
Dialect(Non-Southern):AS	0.523	0.273	1.919	0.057	.
Voice(Midland):Dialect(Non-S):AS	-1.133	0.391	-2.898	0.004	**
Openness	0.327	0.140	2.339	0.021	*
Voice(SouthernAY):openness	-0.888	0.210	-4.237	0.000	***
Dialect(Non-Southern):openness	-0.529	0.248	-2.132	0.035	*
Voice(SouthAY):Dialect(Non-S):openness	0.999	0.337	2.965	0.004	**
Baseline	-0.168	0.155	-1.081	0.282	
Voice(SouthernAY):baseline	-0.469	0.200	-2.337	0.021	*
Voice(SouthernNoAY):baseline	-0.442	0.205	-2.155	0.033	*

Table 4.1: Linear model predicting individual participants' shift from baseline to exposure.

- **Voice*Baseline:** Two-way interaction between Voice and Baseline. Unlike other instances of the predictor Voice, here, Midland is the reference level so that we can answer the important question of whether Baseline influences the degree of shift in the Midland condition

- **Voice*Dialect*SocDes:** Three-way interaction between Voice, Dialect, and SocDes

- **Voice*Dialect*AS:** Three-way interaction between Voice, Dialect, and AS

- **Voice*Dialect*Openness:** Three-way interaction between Voice, Dialect, and Openness

Statistically significant (and borderline significant) comparisons pulled from the full best-fit model are shown in Table 4.1. The non-significant effect of baseline is also included in Table 4.1, since this will be discussed in the results.

Importantly, there is no main effect of Baseline with Midland Condition as the reference level (Est= -0.168 , p=0.282), suggesting that baseline does not influence degree of shift in the Midland Condition. The role of Baseline in degree of shift therefore is not likely attributable to global shifts throughout the course of the experiment, and is instead dependent on the experimental manipulation. Indeed, we see a significant interaction between Condition and Baseline, with those in the Southern

/aɪ/–Present (Est. = -0.469, p = 0.021) and /aɪ/–Absent (Est. = -0.442, p = 0.033) Conditions both showing a more negative correlation between Baseline and Shift. That is, those with higher F1 (more Southern like productions) to begin with, converge less (or even diverge) than those with lower F1 to begin with.

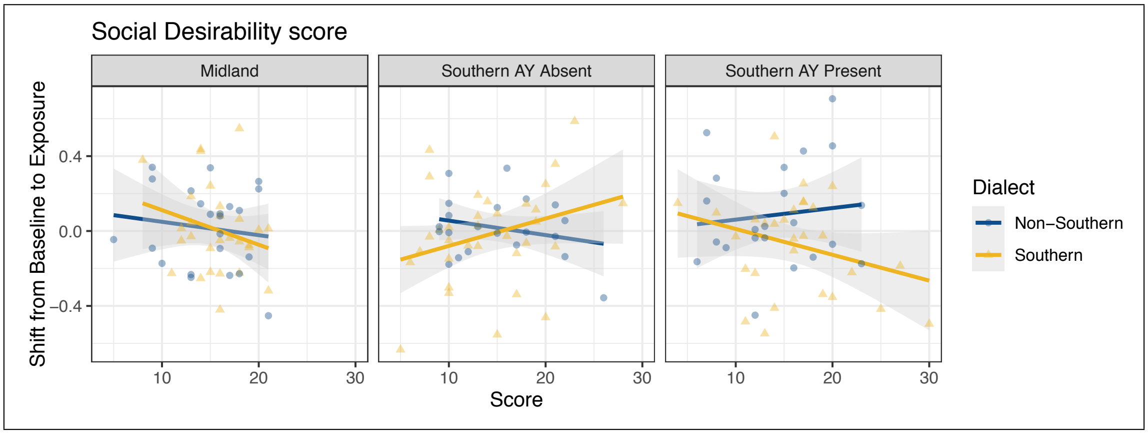

Figure 4.5: Relationship between Marlowe-Crowne Social Desirability Scale score and shift in production. Points indicate participant means.

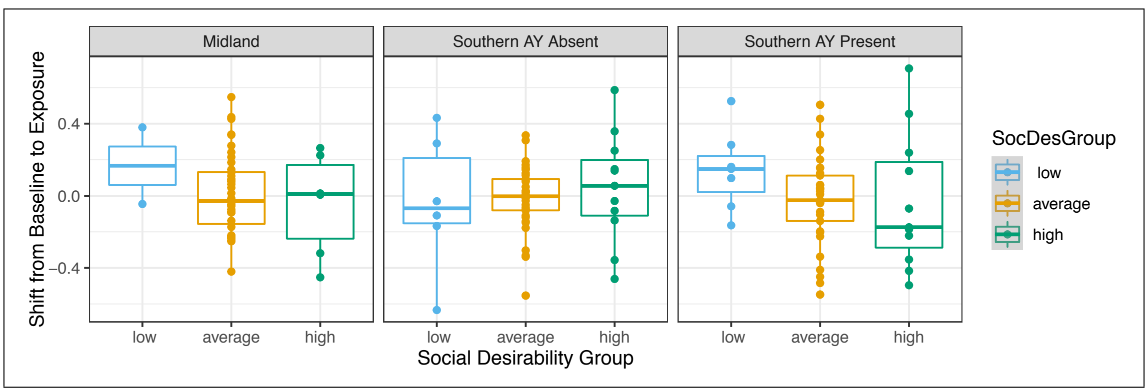

Figure 4.6: Participant shift in production, broken down by Social Desirability Scale score groups proposed in Marlowe-Crowne, and Condition. Points indicate participant means.

There is a main effect of Social Desirability (Est. = 0.364, p = 0.005), suggesting that Southerners in the Southern /aɪ/–absent Condition with higher Social Desirability scores converge more, consistent with the effect we observed for both dialect groups in Experiment 1. Southerners in the Midland and Southern /aɪ/–present conditions show less of an effect of Social Desirability, evidenced by significant interactions between Social Desirability and Condition for the Midland Condition (Est. = -0.746, p = 0.011) and the Southern /aɪ/–Present Condition (Est. = -0.479, p = 0.013), compared to the Southern /aɪ/–Absent Condition (reference level). There is no significant interaction with Dialect. As Figure 4.5 shows, there is a positive correlation between Social Desirability score and degree of shift for Southerners in the Southern /aɪ/–Absent Condition, and a negative trend for those in the Southern /aɪ/–Present and Midland Conditions, with a lack of clear pattern for Non-Southerners. This is consistent with the positive correlation we saw between Social Desirability and degree of shift in the Southern (/aɪ/–absent) condition of Experiment 1, though this effect persisted for both Dialect groups. Figure 4.6 also shows this result, binning scores into the designated categories provided by Crowne and Marlowe (1960) and grouping both Dialect backgrounds together. The negative correlation in the Midland condition may seem surprising at first. However, note that any production shift toward a higher F1 glide would simultaneously be consistent with divergence and hyper-articulation, so that what appears to be divergence may actually be hyper-articulation. It is likely that those with higher Social Desirability scores may shift to hyper-articulate when conversing with a model talker, as a means of being perceived in a more socially desirable way. Lower Social Desirability scorers may undershoot more as the experiment progresses and therefore appear to converge, even in the Midland (control) condition. This would be a (consistent) extension of previous work showing that individuals with higher Social Desirability scores tend to avoid undesirable linguistic behavior, such as interrupting (Natale, 1976). A correlation between Social Desirability and hypo/hyper-articulation would explain the pattern in the Midland Condition, where participants have not heard any speech stimuli that would elicit such shifts. In the Southern /aɪ/–absent Condition, despite this pattern, participants may be more inclined to *converge* when they want to be perceived in a more socially desirable way, and more inclined to diverge when they don't. One reason we may see a positive correlation of convergence and Social desirability in the /aɪ/-absent Condition, but not in the /aɪ/-present Condition is that social motivations may be more important for triggering expectation-driven convergence, particularly if input-driven convergence may be (at least partially) attributable to automatic consequences of perception, or may require less social motivation.

The model also shows significant effects of Openness and Attention Switching scores, which notably are two measures that have been consistently found to correlate with convergence in previous studies. The main effect of AS shows a negative correlation between AS and convergence for Southerners in the Southern /aɪ/–Absent Condition (Est. = -0.457, p = 0.008). That is, those with more "autistic-like" attention-switching capabilities (i.e., greater focus and less attention switching), tend to converge less. This is the opposite of what previous studies have found (e.g., Yu et al., 2013), though these studies have looked at input-driven convergence and suggested that greater AS scores may indicate greater attention to phonetic detail in a convergence task. The model also shows a significant interaction between Voice and AS (Est. = 0.636, p = 0.024), such that those in the Midland Condition are not as influenced by AS—this makes sense, since any shifts in speech in the Midland Condition should not be attributable to convergence and would likely not be predicted by individual differences measures. Further, there is a borderline interaction between Dialect and AS (Est. = 0.523, p = 0.057), suggesting that the affect of AS score on convergence is somewhat less for Non-Southerners in the /aɪ/—Absent Condition. This difference between Southerners and Non-Southerners is clearly illustrated in the middle facet of Figure 4.7. Finally, there is a significant three-way interaction between Voice, Dialect, and AS score (Est. = -1.133, p = 0.004), suggesting a difference in the relationship between Non-Southerners and Southerners in the Midland vs. /aɪ/–Absent Condition. As Figure 4.7 shows, Non-Southerners and Southerners do not behave much differently in regard to the role of AS in convergence in the Midland Condition; however, in the two Southern Conditions, Southerners and Non-Southerners exhibit opposite patterns, such that AS negatively correlates with degree of shift for Southerners, but shows a more positive direction for Non-Southerners. The opposing patterns of Southerners and Non-Southerners in the two Southern Conditions suggests that Attention Switching may be an important part of convergence *or divergence*. Those with lower Attention Switching scores show little shift, while those with higher scores show either greater convergence (i.e., Non-Southerners, who tend to converge in the aggregate) or Divergence (i.e., Southerners, who tend to diverge).

The model also shows a main effect of Openness (Est. = 0.327, p = 0.021), such that higher openness scores indicate greater convergence for Southerners in the /aɪ/–Absent Condition. The positive correlation between Openness and convergence aligns with previous findings (though previous findings looked at input-driven convergence). Openness has been proposed to correlate with convergence for the same reason as Attention Switching—greater openness has been hypothesized to indicate greater engagement with exposure materials and therefore a greater ability for noticing fine phonetic details to converge toward. It is somewhat surprising then, that openness would show a positive correlation in this condition where AS showed a negative correlation, though this may indicate that Openness does not only indicate greater attention to phonetic detail. There is also

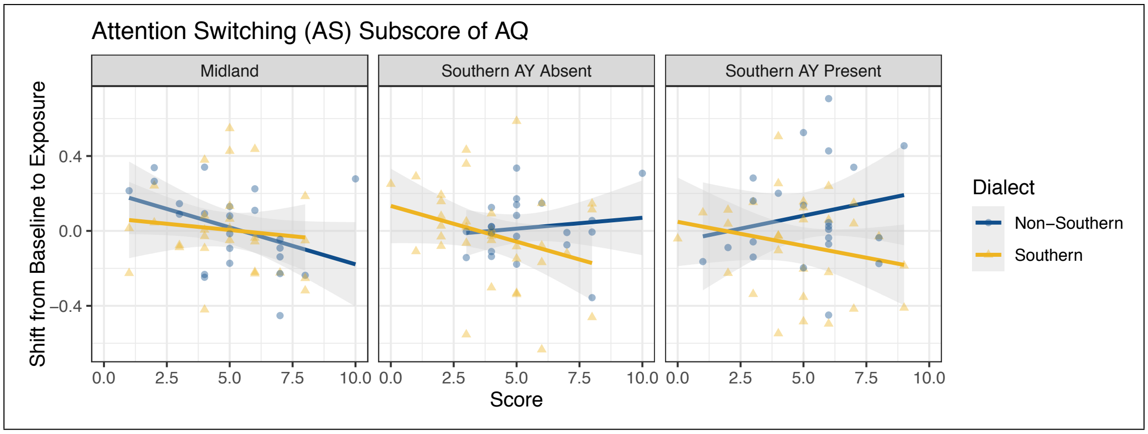

Figure 4.7: Relationship between Attention Switching score on the Autism Spectrum Quotient questionnaire and shift in production. Points indicate participant means.

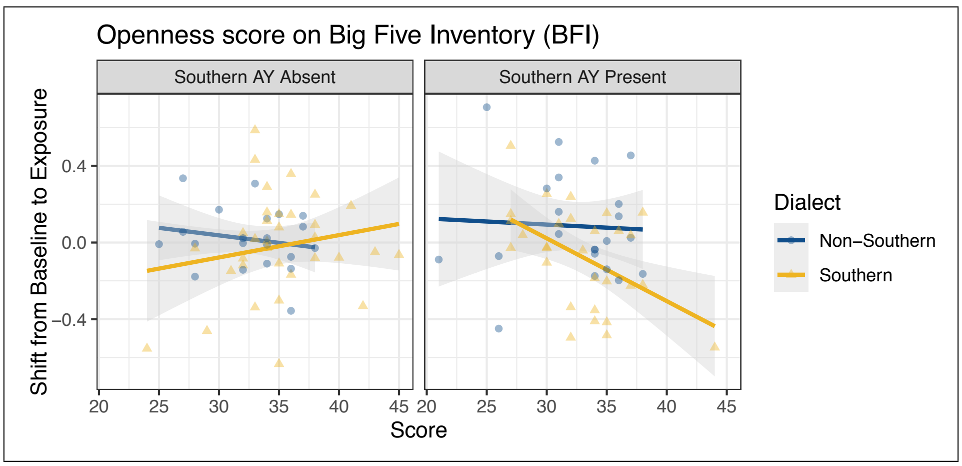

Figure 4.8: Relationship between Attention Switching score on the Autism Spectrum Quotient questionnaire and shift in production. Points indicate participant means.

a significant interaction between Voice and Openness (Est. = -0.888, $p < .001$) suggesting that Southerners show opposite effects of openness in the /aɪ/–Absent and /aɪ/–Present Conditions. As Figure 4.8 shows, Southerners show a positive correlation between Openness and convergence in the /aɪ/–Absent Condition, and a negative correlation in the /aɪ/–Present Condition. The negative correlation in the /aɪ/–Present Condition, which tests input-driven convergence, directly contradicts previous findings showing a positive correlation between Openness and degree of convergence. However, similar to Attention Switching, if Openness is hypothesized to contribute to convergence in that greater Openness scores correlate with greater openness toward observing phonetic stimuli, it makes some sense that again Southerners show this negative correlation. Since here Southerners' aggregate response to stimuli is divergence, it is expected that noticing this stimuli would correlate with an increased response in that direction. When there is no local target available for noticing, as is the case in the /aɪ/–absent Condition, Southerners do not show this same correlation. Further, the opposing patterns for Southerners in the input-driven and expectation-driven conditions may hint toward the possibility that these two types of convergence are not governed by identical cognitive mechanisms or motivations. There is also a significant interaction between Dialect and Openness (Est. = -0.529, p = 0.035), such that Non-Southerners show less of a correlation in the /aɪ/–Absent Condition than Southerners. It is unclear why openness would primarily effect Southerners and not Non-Southerners, who do not appear to show an effect of Openness in either condition. Finally, there is a three-way interaction between Voice, Dialect, and Openness (Est. = 0.999, p = 0.004), suggesting that the relationship between Southerners and Non-Southerners differs across the two Southern voice conditions, though as 4.8 shows, this seems to be driven primarily by Southerners' opposite patterns in the two conditions.

Finally, the model also shows an effect of gender. There is a borderline main effect of gender, suggesting that females tend to diverge in the /aɪ/–Absent Condition (Est. = -0.218, p = 0.052), particularly when compared to the Midland Condition, where females show the opposite pattern (Est. = 0.636, $p < .001$). This is consistent with previously literature that males converge more, though a gender effect was not observed in Experiment 1.

All in all, while some of the individual differences measures reported here are consistent with previous findings, the data is somewhat messy and involves multiple interactions across conditions and dialects. For some measures, Non-Southerners (who tend to converge) and Southerners (who tend to diverge) behave oppositely, suggesting that the measure may be picking up on traits that influence a tendency to *shift* but not necessarily converge. Other measures show opposite patterns for the /aɪ/–present and /aɪ/–absent conditions, suggesting that expectation-driven convergence and input-driven convergence may stem from different motivations or mechanisms. Still, investigation of these individual differences predictors is largely exploratory, and it is difficult to draw absolute conclusions from these results.

4.3.2 Perception

For the perception task, a mixed effects logistic regression model predicting word responses (1) or non-word responses (0) was fit to the data. Responses to non-word stimuli are not of interest and were excluded from the analysis for simplicity and ease of interpretation. All fixed and random predictors were included based on likelihood ratio test significance ($p < .05$). Fixed predictors that significantly improve the model are as follows:

- **Type:** Categorical predictor with levels Target (reference level) and Filler, treatment coded.
- **Condition:** Categorical predictor referring to the voice heard in the exposure phase of the experiment, with levels Southern Ay-Absent (reference level), Southern Ay-Present, and Midland, treatment coded.
- **Item:** Continuous predictor referring to order of presentation of each item, centered.
- **Frequency:** Lexical frequency for each item, using the Log10CD measure from the SUBTLEX Corpus.
- **Item*Condition:** A two-way interaction between Item and Condition
- **Type*Condition:** A two-way interaction between Type and Condition. Since both predictors are treatment coded and put into the model as an interaction, the main effect of Type considers only the Southern Ay-Absent Condition, and the main effect of Condition considers only Target word-type.

The model also includes significant random by-participant slopes for Type and intercepts for Item. The final model is shown in Table 4.2.

Figure 4.9 shows the proportion of "word" responses for each stimulus type, across the three experimental conditions. Bar height indicates mean proportion of "word" responses, with error bars indicating 95% confidence intervals for all items. The points represent individual participant means. In the Midland Condition, /aɪ/ target words (e.g., "bribe" spoken with a glide-weakened /aɪ/ so it sounds like a potential non-word "brahb") are recognized as words the majority of the time, but word response rates are significantly lower than for filler (non-/aɪ/) items. In the Southern /aɪ/–Absent Condition, however, filler words are recognized as words at increased rates, even after hearing a Southern talker who never uses the /aɪ/ vowel. The difference between the Midland condition and the Southern /aɪ/–Absent Condition is statistically significant. There is a main effect of Condition, suggesting that target items are less likely to be recognized as words in the Midland Condition than

Random Effects:	Groups	Name	Variance	Std.Dev	Corr
	Time	(Intercept)	1.872	1.368	
		TypeFiller	1.066	1.033	-0.89
	word	(Intercept)	1.227	1.108	
	Estimate	Std. Error	z-value	p-value	
(Intercept)	2.283597	0.504126	4.53	5.90E-06	***
Type(Filler)	-0.166125	0.297018	-0.559	0.575951	
Condition(Midland)	-0.77721	0.271945	-2.858	0.004264	**
Condition(Southern-AyPresent)	0.567004	0.275731	2.056	0.039747	*
Item	0.002656	0.002982	0.891	0.37312	
Frequency	0.524928	0.156938	3.345	0.000823	***
Type(Filler):Condition(Midland)	0.871558	0.245719	3.547	0.00039	***
Type(Filler):Condition(Southern-AyPresent)	-0.264068	0.2541	-1.039	0.298698	
Condition(Midland):Item	0.00871	0.001718	5.069	4.01E-07	***
Condition(Southern-AyPresent):Item	-0.005093	0.001814	-2.807	0.004998	**

Table 4.2: Mixed Effects binomial logistic regression model predicting responses on the lexical decision task: word (1) nonword (0).

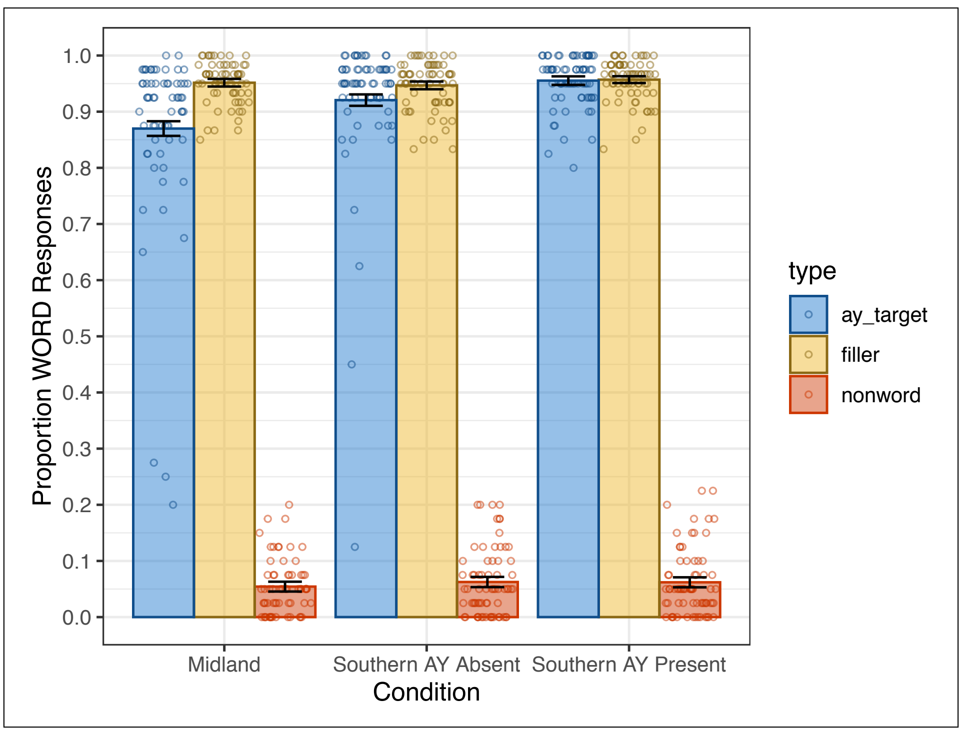

Figure 4.9: Proportion of “word” responses on the lexical decision task across all items with 95% confidence intervals, broken down by Condition and Stimulus Type. Points represent participant means, but bar height and error bars are based on all items, not participant means.

in the Southern /aɪ/–Absent Condition (Est. = -0.777, p = 0.004). Additionally, there is a significant two-way interaction between Type and Condition, suggesting that the difference between Target items and Fillers is greater in the Midland Condition than in the Southern /aɪ/-Absent Condition

(Est. = 0.872, p < .001). This means that, even when participants never hear the /aɪ/ vowel from the model talker, simply hearing Southern speech is enough to shift participants' perceptual /aɪ/ boundaries to encompass more glide-weakened tokens.

In the Southern /aɪ/–Present Condition, target items are just as likely to be recognized as "words" as filler items. There is little difference between the two Southern Conditions; though there is a main effect of Condition, suggesting that Target items are somewhat less likely to be recognized as words in the /aɪ/–Absent Condition than in the /aɪ/–Present Condition (Est. = 0.567, p = 0.04), there is no significant interaction between Type and Condition (Est. = -0.264, p = 0.299). Also note that, in the /aɪ/–Present condition, all participants accept the target words at high rates. Unlike in the other two conditions, no participant has less than an 80% likelihood of identifying a target word as a real word, suggesting that the perceptual shift in response to glide-weakened /aɪ/ is more widespread, and perhaps an automatic consequence of perception. Finally, non-words are recognized as words at consistently low rates across conditions.

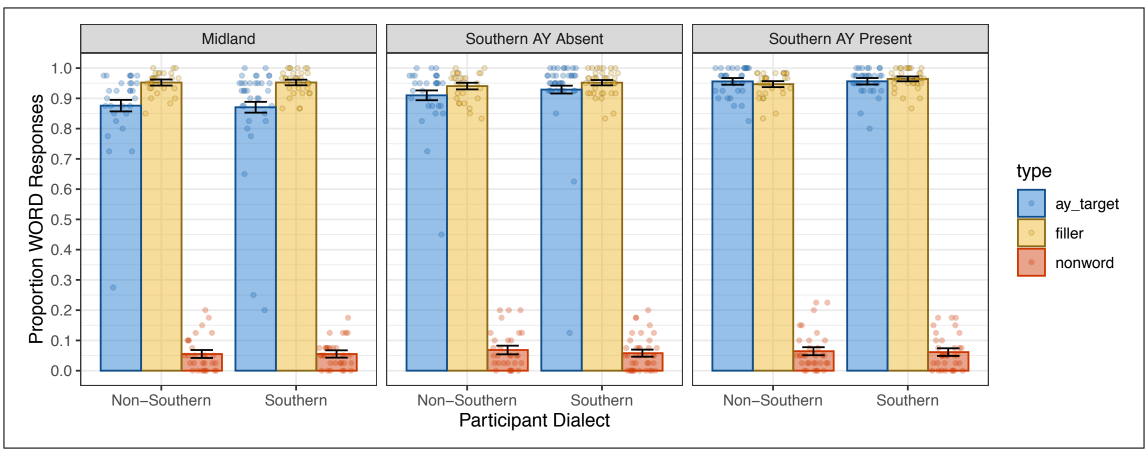

Figure 4.10: Proportion of "word" responses on the lexical decision task across all items with 95% confidence intervals, broken down by Condition, Stimulus Type, and participant Dialect. Points represent participant means, but bar height and error bars are based on all items, not participant means.

Based on earlier findings that Southerners and Non-Southerners behave somewhat differently in production, it is reasonable to expect that individuals from different dialect backgrounds would behave differently in perception as well. This does not appear to be the case here. As Figure 4.10 shows, both dialect groups behave similarly across Conditions. Note that, even in the Midland Condition, Southerners do not exhibit increased word identification rates for target items, suggesting that previous experience with the Southern dialect (or even having a Southern dialect) does not automatically lead to increased word identification rates for glide-weakened /aɪ/-words. Rather, participants must have some reason to expect /aɪ/ words, such as having recently heard a Southern talker. Another way of thinking of this is that Southerners do not default to assuming a speaker is also a Southerner before obtaining sufficient evidence to the speaker's dialect. In the Southern /aɪ/–Absent Condition, Southerners appear to have slightly increased rates of word responses for target items, though this effect is not significant. In fact, including Dialect does not significantly improve the model, regardless of whether it is as a main effect, in a two-way interaction with Type or Condition, or a three-way interaction with Type *and* Condition.[1]

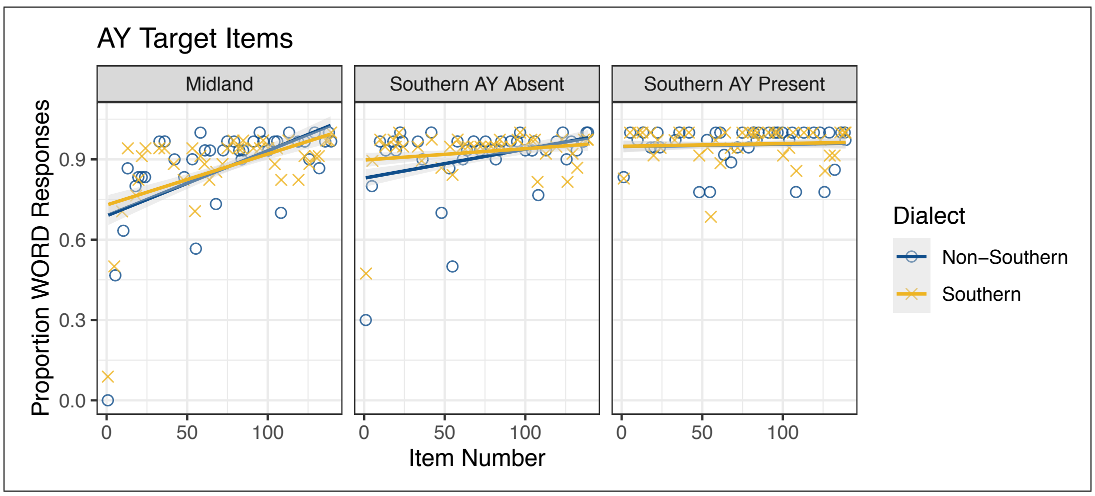

Figure 4.11: Proportion of word responses to target items on the lexical decision task as a function of item order. Point shape and color indicate participant dialect, and individual points represent item means.

Finally, the model shows a significant interaction between Item order and Condition. As Figure 4.11 shows, those in the Midland Condition increase their acceptance of target items as words as the experiment progresses, suggesting that they learn something about Southern speech within the context of the Lexical Decision task alone. Those in the Southern /aɪ/–Absent Condition show a similar tendency to learn across the course of the experiment, though less than in the Midland Condition, evidenced by an interaction between the Midland and Southern /aɪ/–Absent Conditions and Item order (Est = 0.0087, $p < .001$). Further, the influence of Item order is greater in the Southern /aɪ/–Absent Condition than in the Southern /aɪ/–Present Condition (Est = -0.0051, p =

[1]Even with dialect included in the model in a three-way interaction with Type and Condition, the model suggests no significant difference between Non-Southerners and Southerners in the Southern /aɪ/–Absent Condition.

0.005), suggesting that participants in the /aɪ/–Present Condition do not learn anything new about the Southern talker in the lexical decision task that they did not already get from the exposure phase.

Individual differences in perception

The role of individual differences in perception was investigated using a method parallel to that in the production component. A single normalized perception measurement per participant was calculated by dividing a participant's "word" response rate for target words by their "word" response rate for filler words. Rather than simply taking the "word" response rate for target items alone, this measure controls for the fact that participants may have different tendencies for classifying items as "words" across the board, as well as different accuracy rates overall. A value of 1 indicates that target items are recognized at words at the same rate as fillers, while values higher than 1 indicate target items as recognized at words at higher rates than fillers, and values lower than 1 indicate target items are recognized at lower rates than fillers. A linear regression model was fit to the data, predicting this Perceptual Shift measure. Equivalent to the procedure for the production data, a maximal model was fit with a series of three-way interactions between Condition, Dialect, and each individual difference measure, as well as Gender. However, no main effects or interactions came out as statistically significant.

It is unclear why some individual difference measures predict production behavior—several of which have also been found to predict degree of convergence in prior studies—while no measures predict perception. One reason may be that the individual differences measures investigated are not relevant for perception (though this would be surprising since the individual differences literature has tended to show more robust correlations of individual difference measures with perception than with production). This would mean that these measures are picking up on something other than what has been previously proposed (i.e., Attention Switching and Openness measures have been proposed to correspond to noticing of phonetic details, which would certainly be relevant to perception). Perhaps the relevance of these measures is not as straightforward and really has to do with what happens afterwards, when perception translates into production. Another explanation is simply that the findings in the production data are spurious and unreliable. We might expect a similar number of spurious correlations in the perception task, though there is arguably greater variation in production, particularly in regard to the measures used here, such that false correlations might be more likely with production data. A final, perhaps most plausible, reason is that several of the individual difference measures previously found to predict convergence behavior are specific to input-driven convergence (i.e., they involve noticing phonetic stimuli), and the participants perform at ceiling in the /aɪ/–Present Condition, which measures input-driven convergence. The overall lack of perceptual variation may obscure any individual differences measures of perception.

Perception-Production Relationships

Finally, this section compares perception behavior to production behavior on an individual basis for the 159 participants who completed both the production and perception components of the experiment. Figure 4.12 plots the relationship between production and perception using the single-point per individual measures described above. Production shift refers to the difference in mean normalized F1 at 80% between the exposure and baseline phases of the production task. The Perception measures refers to each individual's proportion of "word" responses for target items, divided by that for filler items. The shapes on the graph represent participant dialect, while the color represents self-reported familiarity with the South. While almost all participants have large perceptual shifts in the Ay-Present Condition, a few have quite small shifts in the Ay-Absent Condition. While there is not much to show statistically regarding the relationship between production and perception, due to high perceptual shift across the board, there are a couple of descriptive points to mention. The first is that those few people with smaller perceptual shifts do not belong to a single dialect group and report vastly different levels of familiarity with the South. In other words, there is no obvious reason why these participants did not shift perceptually. We can also draw one intriguing conclusion based on the distribution of the points in the Ay-Absent Condition: those who show little to no perceptual shift have production shifts close to zero. All of the points in this condition below the 1-mark range between -0.2 and 0.3, which is within the middle range of the production shifts. Further, the two participants with the lowest word endorsement rates have production shifts nearly *at* zero. Importantly, if a perceptual shift is a pre-requisite for a shift in production (whether convergence or divergence), we would expect to see no instances of individuals with small perception shifts but high production shifts. This is indeed what we see, and this finding supports the idea that a perceptual shift is a (perhaps largely automatic) necessary pre-requisite for convergence or divergence, but not a sufficient condition for eliciting production shifts. This is an intriguing possibility worth examining in future work.

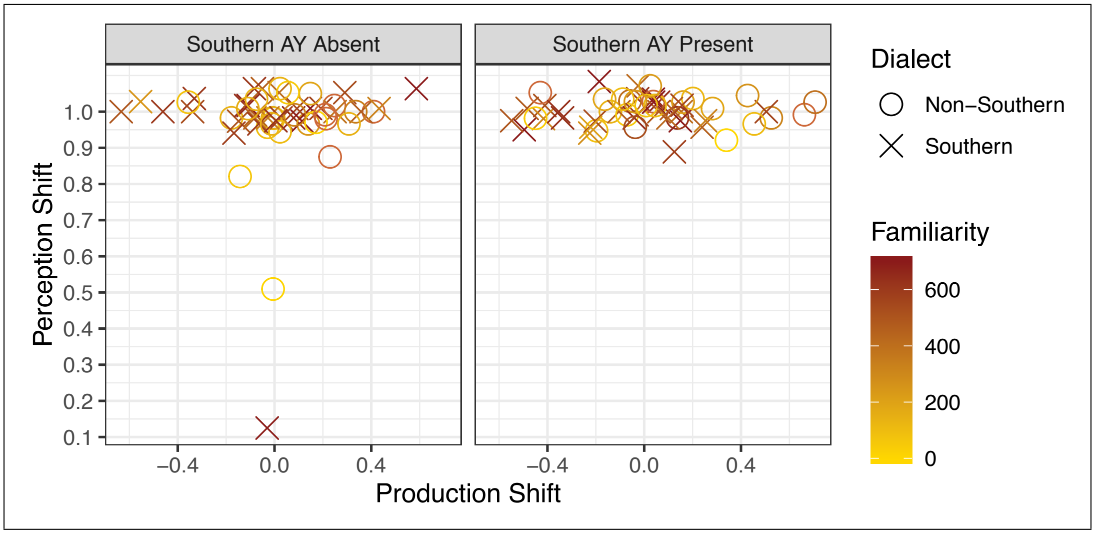

Figure 4.12: The relationship between production shift (measured as the difference in a participant's mean normalized F1 at the glide (80%) in the exposure and baseline phases) and perception shift (measured as an individual participant's "word" response rate for target items divided by that for filler items. Points represent participant means, shapes indicate participant dialect, and color indicates self-reported familiarity with the South.

4.3.3 Discussion

To summarize, Experiment 2 used a combined perception-production individual differences design to compare shifts toward both observed and expected linguistic behavior, as well as to understand the perceptual correlates of linguistic convergence. The perception task used an auditory lexical decision design to measure perceptual shifts in participants' /aɪ/ category boundary to encompass more glide-weakened tokens after exposure to Southern speech. Results show that similar perceptual shifts occur in response to both observed and expected glide-weakened /aɪ/. Crucially, even those who never heard any /aɪ/ tokens from the Southern model talker shifted their perceptual boundaries, suggesting that expectation-driven perceptual learning does occur. The production task used a Word Naming Game speech elicitation paradigm and found a high degree of variation in production responses to the Southern model talker—including convergence, divergence, and speech maintenance, which were largely predicted by participants' baseline productions of /aɪ/. Southerners (many of whom were already somewhat Southern-shifted) actually tended toward divergence, while many Non-Southerners converged. Comparison of individual difference predictors of production behaviors suggest that convergence and divergence may pattern similarly (but differently from maintenance) and further that input-driven and expectation-driven convergence pattern differently in a number of ways. While perceptual shifts did not correlate precisely with shifts in production, there is some evidence to suggest that perceptual responses are necessary (but not sufficient) in inducing shifts in production. The implications of these findings for models of speech perception and production, the perception-production interface, and the relationship between input-driven and expectation-driven behaviors are discussed in detail here.

Perception

The key finding that expectation-driven perceptual learning occurs suggests that perceptual responses do not require direct (local) observation, which is consistent with the findings of Weatherholtz (2015) who found that participants exposed to only part of a novel chain-shift generalized to the rest of the vowel space and exhibited perceptual learning to unlearned parts of the chain shift. Further, expectation-driven perceptual learning has implications for theories of linguistic perception. For one, models of speech perception that rely heavily on statistical learning of associative patterns are well-suited to account for participants' perceptual behavior, however the nature of the units required for such associative processes remains unclear. Are people able to form expectations for glide-weakened /aɪ/ because they recognize statistical co-occurrence patterns (e.g., fronted back vowels and Southern-shifted front vowels tend to occur alongside glide-weakened /aɪ/)? Are they able to extrapolate phonetic values based on structural knowledge (e.g., hearing a lowered /e/ might lead to an expectation for /aɪ/ to be produced elsewhere in the vowel space to avoid overlap)? Do people rely on (perhaps social) categories to group co-occurring patterns (e.g., hearing other Southern-Shifted features leads to expectations for fronted back vowels, crucially because both are recognized as "Southern)?

At this point we can posit at least two distinct associative mechanisms to account for this finding—one relying on top-down knowledge of social category labels and the other relying on bottom-up structural knowledge of purely linguistic patterns. Under the top-down account, participants would crucially recognize the model talker as "Southern," (which we assume they should be able to do based on findings that participants are relatively good at recognizing dialects within their own language (Clopper and Pisoni, 2005; Purnell et al., 1999; Clopper and Pisoni, 2007) and activation of this social category would similarly activate associated linguistic forms, such as /aɪ/-glide weakening, which is a socially salient and recognizable feature of Southern English. Alternatively, under a bottom-up account where convergence is triggered by acoustic cues, participants would rely on distributional knowledge that /aɪ/-glide weakening frequently occurs alongside fronted back vowels and Southern-shifted front vowels, and this knowledge may be utilized to predict /aɪ/-glide weakening from the model talker. Another possibility relying solely on bottom-up acoustic cues is that listeners may posit the location of one vowel in the vowel space based on the location of others. For instance, if /e/ has encroached on the territory of /aɪ/, listeners might expect /aɪ/ to be produced elsewhere in the vowel space. Regardless, under either bottom-up explanation, recognition of the talker as "Southern" is unnecessary. While it is difficult to tease apart the role of bottom-up and top-down mechanisms in the shifts observed in the present experiment, in part because it is impossible to control whether a participant activates social category labels during perception, Experiment 3 begins to tease apart these accounts using a dialect-label mismatch task using the same Word Naming Game design.

Comparison of input-driven and expectation-driven perceptual learning may also speak to the automaticity of perceptual responses to observed behavior, as well as the relationship between expectation-driven and input-driven shifts. Participants in the /aɪ/–present condition show across the board high acceptance rates of target items as words, with no individual word-endorsement rates less than 80%. While many participants had similarly high word endorsement rates in the /aɪ/–absent condition, there were also some participants with quite low word-endorsement rates, suggesting a lack of universal shift toward expected behavior. While the number of participants with very low word endorsement rates for target words in the /aɪ/–absent condition are quite small, we might begin to speculate that the universally high word-endorsement rates in the /aɪ/–present condition indicate more obligatory shifts in response to observed behavior, while similar shifts may or may not occur toward expectations for a number of reasons. One reason for failure to shift might be failure to generate expectations based on observed stimuli, or lack of sufficient previous experience linking observed stimuli to (what would be) an expected variant in the first place (though familiarity with the South and dialect background did not seem to influence whether a participants exhibited perceptual shift). Task effects are also possible; that is, an individual may fail to endorse target items as words despite recognizing that the model talker would exhibit such pronunciation due to demand characteristics or attitudes towards non-standard speech. However, the fact that we don't see any low word endorsement rates in the /aɪ/–present condition suggests that the few low word endorsement rates we do see in the other two conditions are likely not just an artifact of the task. A promising area of further inquiry is in the factors that contribute to whether shifts occur toward expectations and/or the conditions under which expectations are formed in real-time in the first place.

We can also draw some conclusions about the role of dialect background in perception. Southerners don't have higher word-endorsement rates for target items in either Southern condition, suggesting they do not exhibit perceptual learning of Southern glide-weakened /aɪ/ to a greater extent than Non-Southerners. Dialect did not improve the model, and even when added to the model did not reveal any significant differences between dialect groups. This suggests that the smaller convergence shifts observed from Non-Southerners in Experiment 1 were likely not a result of Non-Southerners' failure to posit expectations for glide-weakened /aɪ/. Instead, Non-Southerners may have converged less in production due to other factors such as having less experience producing glide-weakened /aɪ/ or having less reason to signal affiliation with the Southern model talker. It's also possible that Non-Southerners were less able to recall the precise phonetic details of glide-weakened /aɪ/ during the production task when no local target was present, which wouldn't have hindered perception since participants simply categorized instances of glide-weakened /aɪ/ without needing to posit a precise phonetic target from memory. Of course, the difference between Southerners and Non-Southerners in Experiment 1 is simply a descriptive observation, and the role of Dialect was not significant in the model. The lack of influence of dialect background here in the perception data may then indicate that Southerners and Non-Southerners have quite similar tendencies to form expectations about glide-weakened /aɪ/ after all.

The lack of Dialect differences in the perception task also indicates that Southerners do not have higher *baseline* word-endorsement endorsement rates for target items, since they endorse target items as words at the same rate as Non-Southerners in the Midland (control) Condition. This is initially surprising due to the vast literature showing that listeners are influenced by their linguistic background in all kinds of perceptual tasks, ranging from word or segment classifications (Fridland and Kendall, 2012; Iverson et al., 2003; Cebrian, 2006; Wade, 2017b), to semantic priming (Sumner and Samuel, 2009; Warren et al., 2003), or neural activity measured by event related potentials

(ERPs) (Lipski et al., 2012; Conrey et al., 2005). One possibility is that, both groups of participants had higher baseline rates for /aɪ/ word endorsement than they would have for another variant because /aɪ/-glide weakening is a particularly salient feature that either group may have experience with, regardless of dialect background. Support for this explanation comes from the observation that both groups recognize target items as words at greater than chance levels (closer to 85%) in the Midland Condition, suggesting relatively high acceptance of these items as words even in the absence of relevant exposure. Still, it is interesting to note that Southerners who did not previously hear a Southerner do not assume the lexical decision talker is Southern at a higher rate than Non-Southern participants, suggesting expectations are not formed without sufficient evidence and are not based on one's dialect alone.

Based on the perception data, we can also draw some conclusions about the nature of learning linguistic variation. Participants in the Midland Condition appeared to learn from the Southern talker just within the lexical decision task and increased word response rates as the experiment progressed. This also occurred in the Southern /aɪ/–absent condition, but to a lesser extent. However, in the /aɪ/–present condition, no such shift occurred; participants had high word endorsement rates from the beginning of the task, suggesting gradation in the amount already learned vs. left to learn about the model talker. In the Midland Condition, participants had learned essentially nothing about the Southern talker by the start of the lexical decision task, and therefore had a greater opportunity for learning. In the Southern-/aɪ/ present condition, participants did not appear to learn anything more about the Southern talker after the start of the lexical decision task than they had already learned. The pattern shown in the /aɪ/–absent condition, midway between these two extremes, could also be indicative of bi-modal behavior in participants: some had learned a lot in the exposure phase and had clearly-generated expectations by the start of the lexical decision task, while others had not.

Production

The large range of individual production responses to the model talker was unexpected based on the findings of Experiment 1, but ultimately useful for our analysis. Observing a larger number of divergers in this task allowed for examination of individual differences correlates across the entire range of production responses, which we would not have been able to accomplish if participants had primarily converged as in Experiment 1. While methodologically useful for the present task, the different aggregate findings, presumably stemming primarily from differences in baseline productions across the Southern subject populations in Experiment 1 and Experiment 2, brings up important methodological cautions worth mentioning briefly here. For one, pooling subjects from the same source (in this case the Prolific online platform) will not always yield entirely comparable participant groups, even when sampled at random, so it is important to collect and compare relevant information across participant groups to determine the generalizability of findings. Further, caution should generally be exercised when comparing different groups in between-subjects designs. Still, I feel confident that the experimental manipulation is what triggered these production shifts, since as we've seen in the pilot experiment, Experiment 1, and now Experiment 2, those in the Midland condition show no shifts across the course of the experiment, while only those in the Southern voice conditions shift at all (albeit in different directions in Experiment 2), suggesting that that the production shifts we do see are in fact due to the experimental manipulation. Further, comparing *relative* degree of shift across participants, which is the intended goal of Experiment 2, is still useful and reliable regardless of the aggregate results.

It is noteworthy that, of the many individual differences measures examined here in Experiment 2, those that significantly predicted convergence behavior tended to be those measures already noted in the literature. For one, Social desirability positively correlated with degree of convergence for Southerners in the Southern /aɪ/–absent Condition, consistent with the findings of Natale (1975a) and Natale (1975b), who found that scores on the social desirability scale positively correlated with degree of convergence in vocal intensity and in temporal aspects like pause duration. This also corresponds with the positive correlation observed for both dialect groups in Experiment 1. However, the effect was opposite for Southerners in the Southern /aɪ/–present Condition, which tests input-driven convergence like Natale's work. One reason for this may be that social desirability has a greater influence for expectation-driven convergence. If input-driven convergence stems (at least partially) from automatic perception-production links, little social motivation would be required for at least some convergence to occur. Social motivations could conceivably be more relevant for expectation-driven convergence, in which an automatic link between trigger and target is not as straightforward. The potential mechanisms for input-driven and expectation-driven convergence will be discussed in greater detail in Experiment 3.

Another measure found to predict degree of convergence is the Attention Switching (AS) subscore of the AQ. While a negative correlation was observed for Southerners in both Southern voice conditions, a positive correlation was observed from Non-Southerners. This is consistent with previous proposals that AQ positively correlates with degree of convergence because greater Attention Switching also means an increased capability for noticing fine phonetic details to converge toward (Yu et al., 2013). Non-Southerners, who trended toward convergence in the aggregate show little shift at lower AS scores but higher convergence shifts at higher AS scores, suggesting that Attention Switching does indeed facilitate convergence. Southerners, however, who tended to diverge in the aggregate, show greater *divergence* at higher AS scores (and closer to zero shift at lower AS scores), suggesting that noticing of fine phonetic details not only facilitates convergence but also divergence.

Finally, the Openness score on the BFI has been previously shown to correlate positively with degree of convergence. It has been proposed that those with higher Openness scores are more open to observing phonetic details, similar to those with high AS scores. Indeed, we see that Southerners with greater shifts (albeit *away* from the model talker) tend to have higher Openness scores in the Southern Ay-Present Condition, indicated by the negative correlation between Openness and shift in this condition. This makes sense if noticing of glide-weakened /aɪ/ is necessary in order for Southern participants to diverge. In the /aɪ/–Absent Condition, on the other hand, we see a moderate *positive* correlation between Openness scores and degree of shift for Southerners, which was notably not found in Experiment 1. In fact, openness trended toward a negative correlation with convergence in Experiment 1. (This discrepancy may indicate a sporadic finding that does not replicate well, or it may be picking up on the different aggregate behaviors of Southerners who tended to converge in Experiment 1 but diverge in Experiment 2). It is unclear why Southerners' Openness scores would *positively* correlate with shift in the /aɪ/–Absent Condition and negatively correlate in the /aɪ/–Present Condition, but this may indicate somewhat different mechanisms guiding expectation-driven and input-driven convergence. Since this is just one measure that oppositely predicts the two types of convergence, however, and since these correlations hold only for Southerners, more work on this topic is needed to draw any firm conclusions.

The divergence observed in Experiment 2 was somewhat surprising, but may offer insights into some of the factors that may mediate production shifts. First it is worth noting that the divergence observed here seems to be a real response to the experimental stimuli, rather than noise or some other task-based effect. The fact that the AS measure predicted degree of shift oppositely for Non-

Southerners (who tended to converge) and Southerners (who tended to diverge), and the similar perceptual responses for divergers and convergers (i.e., divergers in production didn't diverge in perception), suggests that divergence is a production response that depends on perceptual integration of the variant of interest. However, why Southerners here tended to diverge in the first place remains unclear. There are several possible explanations for divergence. One explanation is attitudinal. While none of our attitudinal measures gauging likability/relatability of the talker or likability of the Southern dialect seemed to predict degree of convergence or divergence, it is possible that divergers as a whole had more negative attitudes toward the Southern model talker or the Southern dialect that were not measured by the post-task survey. For instance, participants did not rate the Southern talker or Dialect on factors such as standardness or clarity, which may come into play in a word-guessing game. It's possible that divergers were in fact those who wanted to sound more "standard" in the context of the word-naming game, particularly in the exposure phase when they had an "interlocutor," and those who perceived Southern speech as less suitable for this purpose may have shifted away from the model talker. Already Southern-shifted participants may be hyper-aware of the negative stigmas associated with Southern speech, particularly when trying to sound "standard" or "clear," and may therefore diverge the most.

Other explanations for divergence are connected to participants' baseline productions. Both Experiment 1 and Experiment 2 found that those more different from the model talker tended to converge more than those already somewhat Southern-shifted (many of whom diverged). Since a greater proportion of the Southerners in this iteration of the experiment were already Southern-shifted, this may explain why *divergence* was the aggregate pattern observed from Southerners. Why, though, would baseline productions influence degree of shift in this way? It does not appear to be the result of global reduction processes throughout the course of the experiment, since such a baseline effect is not observed in the Midland (control) Condition. One explanation may have to do with variant salience. It has often been observed that convergence occurs to salient features but not to less salient features. For instance, Zellou et al. (2013) found that reduced nasality was imitated for longer than increased nasality, which they explained as being a result of the greater novelty and perceptual salience of reduced nasality. Similarly, Podlipský and Šimáčková (2015) argued that imitation of extended (but not reduced) pre-voicing in Czech was a result of the greater salience of extended pre-voicing. While salience may at first be a likely explanation—variants more acoustically similar to one's own are more likely to be noticed in the first place—there is no evidence for this in perception; Southerners or more Southern-shifted participants did not show *less* perceptual learning (or perceptual "unlearning"/divergence) toward the Southerner, indicating less noticing of this variant. Rather, Southerners and Non-Southerners behaved perceptually the same. The role that participants' baselines play does not therefore seem to be in *noticing* the variant of interest. A different reason already Southern-shifted participants may tend to diverge is a general goal to be similar (but not *too* similar) to an interlocutor. Non-Southern-Shifted individuals may be able to shift a great deal toward the model talker while still not imitating them precisely, while Southern-shifted individuals may have to shift *away* from the model talker in order to accomplish this goal and maintain some degree of social distance. However, we see a gradient effect of baseline on convergence/divergence, not just a drop-off of convergence after a certain point that may be considered too much imitation. (Of course, different individuals may have different ideas of what constitutes *close enough* vs. *too similar* to an interlocutor's speech). The most likely explanation, though, is that divergence among Southerners is a result of attempts at contrast preservation, which has been noted as a reason for inhibition of convergence in the literature. For instance, Nielsen (2011) found that participants imitated lengthened VOT of voiceless stops, but not shortened VOT, because these productions would not be sufficiently distinguishable from their voiced counterparts. Similarly, Kim and Clayards (2019) found that imitation of vowel quality was mediated by attempts at contrast preservation (though vowel duration was not, likely because this is not generally a primary contrastive cue in English). However, note that Mitterer and Ernestus (2008) found results inconsistent with contrast preservation. They observed that Dutch pre-voicing was not imitated and suggested this was because it was not relevant for phonological contrasts (though, non-contrastive features may also be less salient, which could account for these findings). Still, contrast-preservation seems to be a likely explanation for the divergence observed among Southerners. The more glide-weakened an individual is already, the less they have to shift for their production of /aɪ/ to encroach on the territory of /ɑ/ or /æ/. While non-Southerners may be able to converge a great deal without concern for contrast preservation, Southerners (particularly those who already produce somewhat weakened /aɪ/ glides may not be able to shift toward the Southern talker much at all without threatening contrast preservation. In fact, after exposure to the Southern model talker, these individuals may actually hyper-correct and *diverge* from the talker due to heightened attention toward contrast maintenance. This would explain why perceptual shifts appear to be necessary for divergence, as well as why baseline negatively correlates with shift.

The Production-Perception Relationship

While some individual differences measures predicted production, none predicted perception, which may suggest that perception and production are not as clearly linked as previously thought. However, it's likely the case, as Schertz and Clare (2019) note about perception-production relationships, that "...an apparent lack of individual link may erroneously be misinterpreted as a 'complex' relationship between the modalities when there may in fact be a straightforward relationship but with inherent differences between perception and production leading to different sorts of variability across the two modalities that obscure the link" (p. 11). In other words, since the perception and the production tasks utilized here are quite different, they may be more prone to different sources of variability. For one, we see generally high word acceptance rates in the perception task, and these ceiling effects may wash out some individual variability in perception. In contrast, production is generally *more* prone to variability from a variety of sources such as speech errors or fatigue. Further, demand characteristics may differ across the two tasks. While participants seemed entirely unaware of the goal of eliciting convergence in the production task, the goals of the perception task were more transparent, which may influence whether participants accepted target items as words. There are a number of other sources of variability that may differ across the two tasks that may obscure any production-perception links, so it would be hasty to suggest that perception and production are not strongly or straightforwardly tied together.

Though perception and production behavior on an individual level revealed no obvious correlation, we can draw some descriptive conclusions that do support an individual perception-production link. We can generally characterize those with lower word-acceptance rates for target items (indicating little to no perceptual shift) as those who tend to maintain their own speech norms and shift very little in production, whether toward or away from the model talker (i.e., they *maintain* their speech norms). If perceptual shifts are a necessary prerequisite for production shifts, it would not be surprising to find that minimal perceptual shifters also shift little in production. Even finding high shifters in perception who shift little in production is consistent with this claim, because production shifts may fail to result from perception shifts due to social motivations inhibiting convergence, speech errors, or lack of motor practice with the target variant. What we would not expect to find is a production shift for those without a perception shift, and in fact, this isn't attested in the data; *nobody* fails to shift in perception then shifts a lot in production. The qualitative relationships between individuals' perception and production behaviors, then, are consistent with the claim that perceptual shifts are necessary for production shifts, however, the number of minimal shifters in perception are quite low in the present data, so this claim should be tested in future work. If this pattern does hold on a larger scale, it would suggest that perceptual shifts are a necessary (but not sufficient) prerequisite for production shifts, consistent with the interpretation of the apparent lack of perception-production link in Kraljic et al. (2008). It also reinforces the point mentioned above that convergence and divergence may stem from similar perceptual shifts. Future research should continue to look into this promising area of the perceptual correlates of linguistic convergence and divergence.

The relationship between input-driven and expectation-driven convergence

Finally, we can draw some conclusions about the relationship between input-driven and expectation-driven perceptual and production responses, which were directly compared in Experiment 2. There is some evidence that the two processes are not entirely equivalent. For one, they exhibit descriptively different time courses in production. For both dialect groups, shifts in response to Southern /aɪ/–present stimuli continued in the same direction, even into the post-exposure phase, whereas those in the /aɪ/–absent condition returned to their baselines in the post-exposure phase, indicating that expectation-driven convergence may be longer lasting than input-driven shifts. Likewise, differences were observed in perception: input-driven shifts were apparently universal, while some individuals failed to exhibit expectation-driven shifts. Further, while continued learning was possible throughout the course of the lexical decision phase in the Southern /aɪ/–absent condition, such was not the case in the /aɪ/–present condition, though this could reasonably be attributed to ceiling effects. Finally, we see that the same individual differences measures significantly predict both input-driven and expectation-driven production shifts, but in opposite directions at times, such as with the Social Desirability scale measure and the Openness score of the BFI, for Southerners.

One possibility is that both types of convergence tap into the same phonetically-detailed linguistic representations but may get there by different means, which may be what these differences are picking up on. Both ultimately activate a phonetically detailed representation of glide-weakened /aɪ/, but input-driven convergence may activate this representation via activation of the form itself, which may persist longer in memory if it is just heard (explaining the observed longevity) and may be more automatic, such that any perceptual shifts translate directly into production unless external factors (such as social motivations) intervene. On the other hand, expectation-driven convergence may activate detailed representations of glide-weakened /aɪ/ via either social categories or a network of other Southern-shifted vowels. Of course, the latter is dependent on having these links exist in the first place and forming relevant expectations (which may depend on dialect experience and may not be universal or automatic, and thus more susceptible to influences such as social desirability). While it is generally accepted in models such as exemplar theory that external, contextual factors (such as social information) as well as linguistic factors can trigger activation of linguistic forms, these two different triggers have not been expressly compared and perhaps behave differently. The nature of different associative links could be a fruitful direction for better understanding the mental representations of linguistic and social knowledge.

Chapter 5

Experiment 3: Disentangling "Top-Down" and "Bottom-Up" Processes in Expectation-Driven Convergence

This chapter reports the results of a dialect-label manipulation task using the same Word Naming Game design as in the previous experiments. This version of the experiment manipulates both top-down (dialect label) and bottom-up (actual talker dialect) cues to determine their relative influence in convergence. In the Midland-Voice, Southern-Label Condition, participants hear a Midland talker but are told that it is a Southerner. Any shifts toward Southern glide-weakened /aɪ/ in this condition would indicate that expectations can be formed (and utilized as targets of convergence) by top-down information alone. In the Southern-Voice, Midland-Label Condition, participants hear a Southern talker but are told that the talker is from the Midwest. Any shifts toward Southern speech in this condition would indicate that bottom-up information such as the acoustic properties of other Southern-shifted vowels can trigger expectations about glide-weakened /aɪ/. I will refer to the Southern-Voice, Midland-Label Condition as "Voice-Triggered" and the Midland-Voice, Southern-Label Condition as "Label-Triggered," since these are the relevant cues that would trigger shifts toward glide-weakened /aɪ/ in each condition. As before, participants never hear any tokens of

/aɪ/ from the model talker in either the exposure phase or auditory instructions of either condition. Just as in Experiment 2, perception shifts are also investigated after exposure to the two differently labeled model talkers to understand the role of top-down and bottom-up information in perception as well as production.

The specific research questions are as follows:

1. Do participants shift (in perception and/or production) toward glide-weakened /aɪ/ when exposed to a Midland Accented talker who is labeled as "Southern"?
2. Do participants shift (in perception and/or production) toward glide-weakened /aɪ/ when exposed to a Southern Accented talker who is labeled as "Midwest"?
3. Do these behaviors differ by participant dialect background?

5.1 Background

There are several potential explanations well suited to accounting for the phenomenon of expectation-driven convergence observed in Experiment 1, ranging from top-down accounts dependent on recognizing social categories to bottom-up accounts dependent on recognizing co-occurrence patterns among observed and expected variants. This section provides some background on potential mechanisms by which top-down or bottom-up information could trigger expectation-driven phonetic convergence.

5.1.1 Socially-rooted "top-down" mechanisms

There are several existing theories that could reasonably account for expectation-driven convergence resulting from social, "top-down" cues. Two of these, which allow for social triggers and rely on automatic activation of associative links, are Exemplar Theory and Social Priming, both of which will be reviewed here. While exemplar models of speech perception and production rely on tight perception-production feedback loops, which may be well suited to explain input-driven convergence, they may also allow for tight links between social and linguistic information (Pierrehumbert, 2001). Under such models, when perceived linguistic items are mentally stored, they are stored along with not only detailed phonetic information, but also detailed information about the social situation, including information about the interlocutor who uttered the form and the context in which the form was uttered. When primed by concepts relating to social categories (e.g., being primed by the concept of "Australia" as in Hay et al. (2006a) and Hay and Drager (2010)), linguistic exemplars associated with Australia (for instance, words spoken by an Australian or in Australia) will become activated in working memory, and subsequent speech will be weighted more heavily on these activated exemplars. Shifts in speech, then, might emerge "as a subtle automatic consequence of the socioindexical labeling of selected exemplars" (Sanchez et al., 2015, p. 90), without having locally observed the variant to be shifted toward. Linguists have appealed to such models to account for the influence of social information on linguistic processing (e.g., Hay et al., 2006a; Hay and Drager, 2010) as well as addressee- and topic-based style-shifting (e.g., Sanchez et al., 2015; Drager et al., 2010). While such models may be able to explain the phenomenon of expectation-driven convergence, they may require additional steps beyond these automatic shifts to account for divergence or modulation of accommodative behaviors by attitudinal factors.

Social priming is in many ways similar to the "activation" of stored knowledge proposed by exemplar models. While social priming explanations are not utilized often in linguistic research (such mechanisms are appealed to to explain similar behavior in social psychology) they seem similarly well-suited to explain the mental relationship between social and linguistic information. Social priming, or the "nonconscious activation of social knowledge structures," encompasses a wide range of phenomena across "a wide array of psychological systems, such as perception, motivation, behavior, and evaluation" (Bargh, 2006, p.147). While the range of phenomena that could be considered "social priming" is quite large, the classic examples of social priming involve exposing participants to a "prime" related to some social concept, which then activates social knowledge surrounding that concept, which ultimately affects participant behavior or judgment. For instance, individuals who are subliminally primed with concepts relating to groups that help others (e.g., nurses) have been suggested to show more willingness to help in an unrelated task (Aarts et al., 2005) or to pick up a dirty tissue (Custers et al., 2008). Similarly, Bargh et al. (1996) found that, after being exposed to concepts relating to the concept of "elderly," participants subsequently walked down a hallway more slowly. While, to my knowledge, social priming has not been appealed to as a plausible explanation for the influence of social information on linguistic perception or production, such a mechanism could well account for such findings. In particular, an appeal to affective motivations in non-conscious goal pursuit in the social priming literature (Aarts et al., 2008) may be useful in understanding divergence or the modulation of accommodative behaviors by attitudinal factors.

It is important to note, however, that "social priming" research is at the heart of social psychology's recent replication crisis. After a well-known psychology researcher admitted to falsifying data in several published papers, several well-established findings in the field of social psychology also failed to replicate (Carey, 2011; Earp and Tramifow, 2015), shedding doubt on the reliability of social priming research. Though replicability is a current issue for many fields, including medicine, economics, and other fields, and is not limited to psychology or social psychology, it highlights the importance of rigorous methodology and statistical analysis. Still, some findings in linguistics that involve social concepts automatically influencing judgments or behavior (while not specifically deemed "social priming" *have* been at least partially replicated. As noted above, Niedzielski (1999) found that regional labels (Detroit vs. Ontario) influenced classification of a vowel on a synthesized continuum. This finding was replicated in a different lab by Hay et al. (2006a), who found the same effect for different New Zealand and Australian labels applied to the same New Zealand speaker. Further, Hay and Drager (2010) and Drager et al. (2010) went on to find that simply evoking concepts of Australia vs. New Zealand resulted in the same general effect.

Under both an exemplar theoretic or social priming account, the social concept itself is enough to directly influence linguistic behavior, and some studies have indeed found evidence for shifts in speech production without an interlocutor present (Hay and Drager, 2010; Drager et al., 2010; Sanchez et al., 2015). However, it is not clear whether expectation-driven convergence would operate under the same mechanisms. There is some evidence to the contrary. For instance, in Experiment 1, convergence occurred only during the Exposure Phase in which participants were interacting with a model talker, but not in the immediately following Post-Exposure Phase. Further, it has been repeatedly shown that information that is understood as being about a person is structured differently and retained better than information that is not attributed to a speaker, such as unrelated items in a list (e.g., Asch, 1946; Chartrand and Bargh, 1996), suggesting that general concept activation may affect processing differently than if a social category is explicitly assigned to a particular talker.

Other accounts place more emphasis on the importance of having an interlocutor to converge toward, and (unlike exemplar models or social priming accounts) tend to assume some degree of conscious control over accommodative behaviors, such as Bell's Audience Design model of style shifting (Bell, 1984), which proposes that language users design their speech style based on those around them. Similarly, Communication Accommodation Theory (CAT) (Giles et al., 1991), suggests that individual shift their speech toward or away from their conversational partners in order to manage social distance. Both could presumably account for expectation-driven convergence, and while social factors are crucial to both, neither offer an account of the cognitive mechanisms responsible, including how social factors ultimately trigger linguistic expectations for individuals to converge toward.

5.1.2 Structurally-rooted "bottom-up" mechanisms

Under a structurally-rooted, bottom-up account, expectation-driven convergence would be triggered by low-level acoustic cues. For instance, individuals may rely on knowledge of statistical co-occurrence patterns between *observed* variants and an unobserved form that tends to co-occur with that variant, and would posit expectations for the unobserved variant based on this statistical knowledge. In the present case, convergence may be triggered by knowledge that glide-weakened /aɪ/ frequently co-occurs with other Southern-shifted variants that *are* observed from the model talker. There is some evidence that language users learn distributional patterns of variant co-occurrence within a talker. For instance, Theodore and Miller (2010) found that participants developed predictions about a talker's VOT for one voiceless stop based on knowledge of their VOT use for a different voiceless stop.

For such a statistical co-occurrence account to sufficiently explain the present data, it would be necessary to establish that the Southern dialect is in fact a coherent system. That is, do such co-occurrence patterns even exist for listeners to pick up on? "Coherence," a notion introduced by Guy (2013), is essentially the "systematic co-variation of variable features which have some social characteristic in common" (p. 64). However, we cannot assume that coherence actually exists for the Southern dialect. This may seem to be an odd statement; after all, deeming something a "dialect" implies a set of shared features and some degree of uniformity. However, with the recent surge of interest in coherence, there has been some doubt shed on assumptions that even standard languages *are* coherent (e.g., Grondelaers and van Hout, 2016; Guy and Hinskens, 2016), and it's unclear why regional dialects would be any more so. Recent evidence suggests, in fact, that some may not be. Becker (2016), for example, found that over 70% of speakers were incoherent in their use of three non-standard New York City English variants: vocalized r, raised BOUGHT and raised BAD, meaning one speaker's use of one of these variants could not be predicted by their use of another.

Investigating the coherence of the Southern dialect

In order to determine the plausability of a mechanism accessing knowledge of variant co-occurrence, I conducted small study of coherence of eight Southern vowels, using the Raleigh Corpus (Dodsworth). This small investigation tests whether the Southern dialect coheres at the individual level (e.g., whether the acoustic properties of one vowel such as /aɪ/ can be predicted based on the acoustic properties of another vowel). The vowels included in this analysis are as follows:

- The nuclei of BAIT and BEET are lowering and backing in Southern English. While almost all Southern dialects show retraction of BAIT, only those more advanced in the Southern Shift show retraction of BEET. Though Raleigh is not considered to be part of the Inland South, which typically exhibits retraction of BEET, some speakers in the corpus do exhibit "more Southern" productions of this vowel. These vowels are measured as the inverse of the vowels' normalized diagonal measurements (i.e., normalized F2 − normalized F1) at the vowel nucleus, which indicates the degree of retraction. The inverse of this value is used so that all measurements are in the same direction. A higher number therefore indicates a more retracted (i.e., more Southern) vowel.

- The nuclei of BET, BIT, and BAT are raising and fronting in Southern English. BIT only follows this trajectory in the most advanced Southern speakers, particularly in the inland South, though some Raleigh speakers do produce more peripheral BIT nuclei. These vowels are measured as the normalized diagonal at the vowel nucleus. A higher number therefore indicates a more peripheral (higher and fronter) vowel, indicating a more Southern production.

- BITE is monophthongizing in Southern speech. Monophongization is measured as the inverse Euclidean distance between the nucleus and glide in the F1-F2 vowel space. A higher number therefore indicates a more monophthongal (i.e., more Southern) production of BITE.

- BOOT and BOAT are fronting in Southern speech. Fronting of these vowels is measured as the normalized F2 value at the nucleus, such that a higher value indicates a more fronted (i.e., more Southern) vowel.

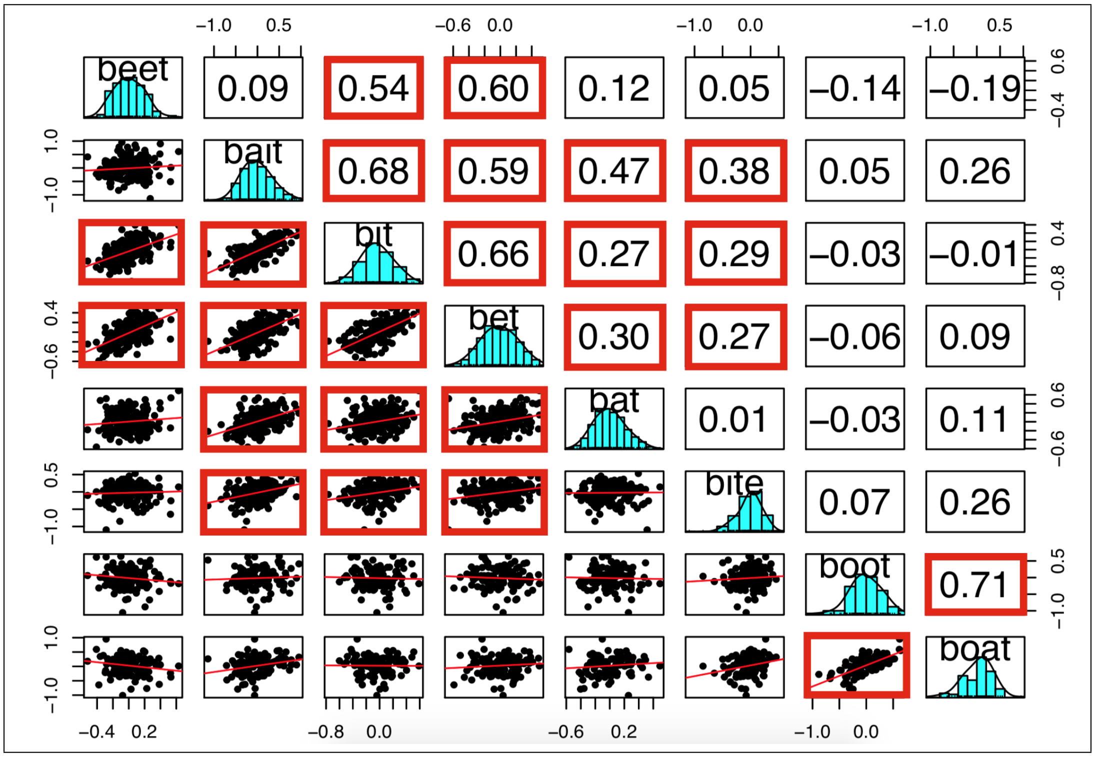

Figure 5.1: Correlation matrix for all vowel classes. Scatter plots are fit with linear regression lines. Numbers indicate Pearson's R. Boxes highlighted in red indicate statistically significant correlations with adjusted p-values of $p < .05$.

Because Southern vowel production is highly correlated with age and social status, we cannot rule out that any covariation on a speaker by speaker basis is actually picking up on a correlation with other social features. For that reason, the values used in the individual coherence analysis are the residuals from a model fit to control for these additional factors. A linear mixed effects regression model was fit to the data using the lme4 package in R (Bates et al., 2015). The relevant measurement outlined above is used as the dependent variable, with participant birth year, occupation category (white or blue collar), and word duration as fixed predictors, and interactions between phone and birth year as well as phone and occupation category. The model includes random intercepts for word but not for speaker. This follows the methodology of Tamminga (2019) in order to control for known sources of variability. For instance, different distributions of lexical items across participants' interview may yield different mean vowel values, and the model attempts to control for such uneven distributions. In addition, the model controls for known social predictors so that we can focus on individual coherence, macro social categories aside.

In all instances, pairwise correlations are used to determine coherence for each pair of vowels. Twenty-eight pairwise correlations were run. Pearson's r values are reported, and significance levels are at the $p < .05$ level. Importantly, when conducting so many pairwise correlations, it is important to correct for multiple comparisons. I use the Hochberg-Benjamini correction, which adjusts for the Family Wise Error Rate (FWER). All p-values reported are adjusted accordingly.

Figure 5.1 displays the results of this investigation. The numbers indicate the Pearson's r for each comparison, and scatter plots are fitted with linear regression lines. The statistically significant (adjusted $p < .05$) are highlighted in red. Results show that the /aɪ/ vowel (BITE) is somewhat predictable based on a speaker's production of three other vowels: BAIT (R = .38), BIT (R = .29), and BET (R = .27). These correlations are not particularly strong, especially when compared to the correlations exhibited between other vowel pairs. However, it is still plausible that participants could converge toward monophthongal /aɪ/ based on knowledge of its co-occurrence with other Southern-shifted vowels, particularly BAIT, BIT and BET.

There is a final potential bottom-up explanation for expectation-driven convergence that does not rely on recognition of statistical co-occurrence patterns and the coherence of the Southern dialect. We can imagine that individuals might also be able to predict a speaker's production of one vowel given the location of other vowels in the vowel space, particularly if these vowels are involved in a chain shift. Weatherholtz (2015), for instance, has demonstrated listeners' ability to make such extrapolations with novel chain shifts. Listeners exposed to /u/, /o/, /ɑ/ /ɔ/ vowels undergoing a novel chain shift (either back-vowel lowered or back-vowel raised shift) formed predictions for /ʊ/ consistent with their training. Though the precise mechanisms behind this behavior are not clear, it could have to do with positing the location of one vowel in the vowel space based on its expected relationship with other (shifted) vowels. It's possible then that, given exposure to other vowels in the Southern-vowel shift, listeners would posit a different realization of /aɪ/. For instance, if /e/ occurs where /aɪ/ might be expected, this may push the boundary between the two phonemes, triggering different predictions for where /aɪ/ would be in the vowel space and allowing for a wider range of pronunciations to be classified as /aɪ/.

One consideration is that different results from different participant backgrounds may hint at which mechanism is better able to account for the data (non-Southern participants with less experience with the Southern dialect should be equally able as Southern participants to extrapolate the position of /aɪ/ based on the location of other vowels in the vowel space, but would not be expected to have as much pre-existing associative co-occurrence information to access).

5.2 Methods

5.2.1 Participants

120 participants were originally recruited to participate in Experiment 3. However, just as with Experiment 2, due to a glitch that resulted in failure to save a portion of the audio recordings, additional participants were recruited so that 120 participants would have audio data, which resulted in more participants having perception data than production data. In total, 146 participants were recruited. Of these, 75 were recruited through Prolific and completed the experiment online for $10 USD. The remaining 71 were recruited through the University of Pennsylvania subject pool and were given course credit for their participation. Of these, 11 participated in the Language Variation and Cognition Lab in the University of Pennsylvania's Linguistics Department, and 60 participated remotely online.

Of the 120 participants with audio data, 5 were excluded for missing more than 20 clues or having poor sound quality that made formant measurements unreliable. 115 participants are therefore analyzed here. Of the 146 total participants, 7 were excluded from the perception analysis for for having less than 75% accuracy on non-target items, resulting in 139 participants left for the perception analysis. As in Experiments 1 and 2, half of participants spent the majority of their school-aged years (5-18) in the U.S. South and were thus labeled as "Southern" (N=75 for perception, N=51 for production), while the remaining half spent this time in the U.S. outside of the Southern isogloss shown in Figure 2.4, and were labeled as "Non-Southern" (N=64 for perception, N=64 for production). Unlike in Experiment 2, Southern participants' baseline productions are not particularly Southern-Shifted, so an unusually high amount of divergence should not be anticipated (Figure 5.2. Participants were all native speakers of American English and reported no speech or hearing impairments.

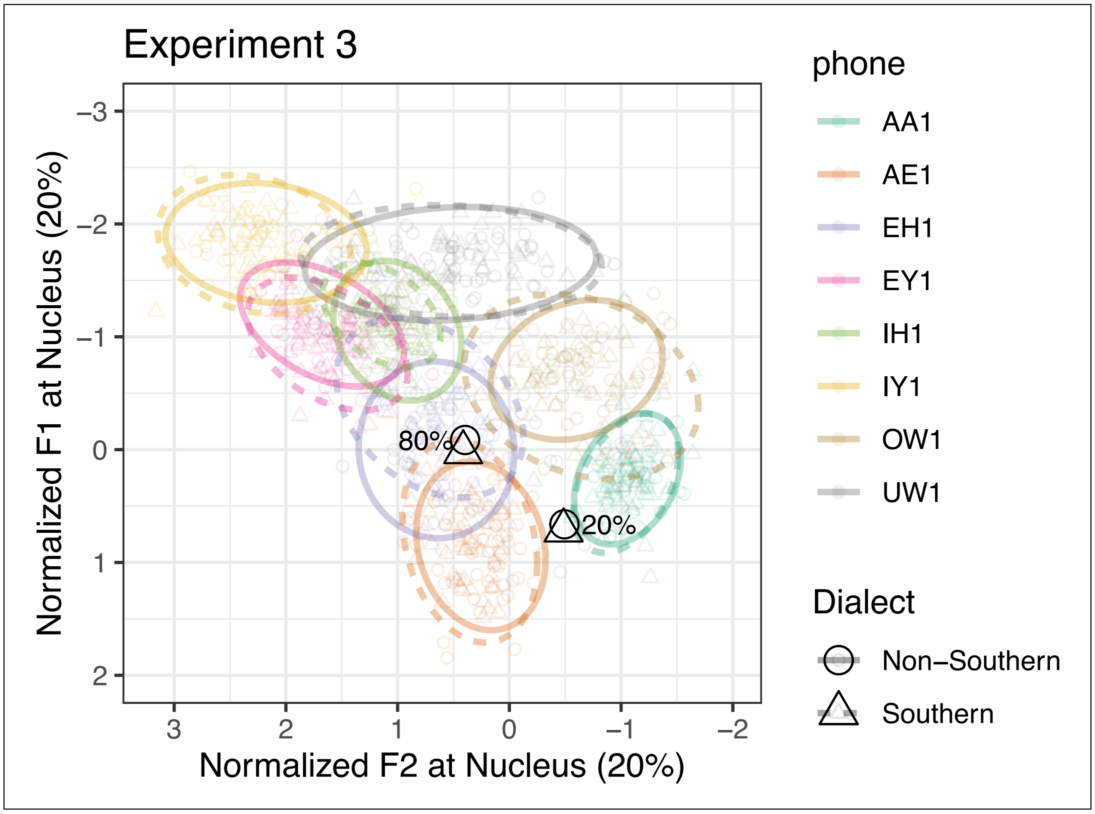

Figure 5.2: Southerner vs. Non-Southerner vowels. Each point represent participant means, and linetype indicates dialect group. Shapes indicate Southerners' (triangle) and Non-Southerners' (circle) productions of the /aɪ/ vowel at the nucleus (20%) and glide (80%).

Of the 115 participants included in the production analysis, 114 disclosed demographic data. Of these, 73 identified as female, 39 as male, and 2 as Non-binary. In regard to maximum educational attainment, 63 reported some college, 14 a Bachelor's degree, 19 a high school diploma, 11 an associates degree, 6 an advanced or graduate degree and 2 other. 68 identified as White, 14 as Black or African American, 12 as Asian or Pacific Islander,10 as Hispanic or Latinx, 3 as Native American or American Indian, and 7 as Other. Mean age is 23.42 with a range of 18–58. Of the 139 participants included in the perception analysis, 137 disclosed demographic data. 87 identified as female, 48 as male, 2 as non-binary. 71 reported some college, 22 obtained a Bachelor's degree, 21 a high school diploma, 14 an Associate's degree, 8 an advanced or graduate degree, and 2 reported "other." The racial breakdown is as follows: 80 White, 17 Black or African American, 13 Asian or Pacific Islander, 15 Hispanic or Latinx, 4 Native American or American Indian and 8 Other. Mean age is 24.68 with a range of 18–58 years.

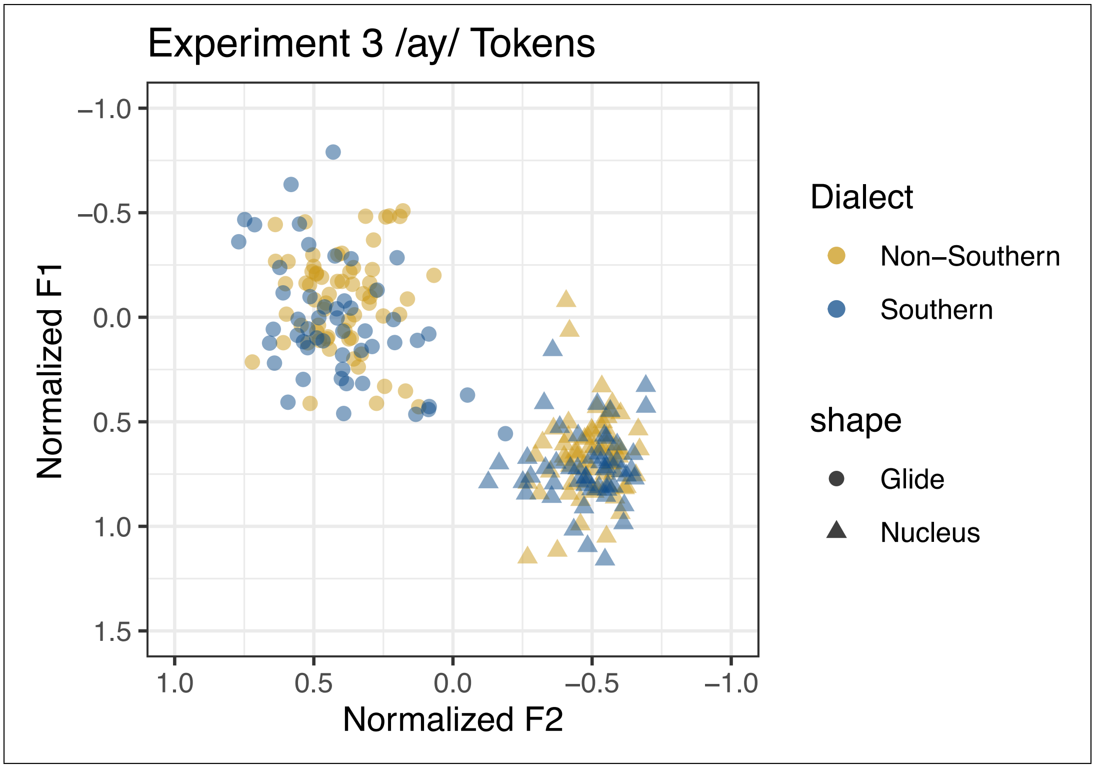

Figure 5.3: Individual speaker means for the /aɪ/ vowel, with triangles indicating the nucleus (20%) and circles indicating the glide (80%).

5.2.2 Procedure

As in Experiments 1 and 2, the experiment was designed using the PennController toolkit (Zehr and Schwarz, 2018) for Ibex and administered via participants' web browsers, and participants were recorded through their computer microphone. The initial demographic questionnaire and follow-up survey were identical to those in Experiments 1 and 2 and will not be repeated here. Please refer to Section 3.2 for these details.

The procedure of the Word Naming Game utilized the same combined production-perception design as in Experiment 2 with several key changes. For one, the two different voice conditions Midland (/aɪ/–absent) and Southern (/aɪ/–absent) had swapped dialect labels, whereas in the previous two experiments no labels were provided at all. At the beginning of the exposure phase, participants were introduced to the model talker they would hear giving clues, who provided short instructions over participants' headphones. In addition to the instructions, the model talkers also introduced themselves as being from a particular dialect region and specifically called attention to their dialect. Rather than assigning the talkers dialect labels in the instructions of the experiment, the model talkers introduced their own dialect labels in their real accents so that there was no chance participants would think that the labels were inaccurate and swapped by mistake. The script the model talkers read is as follows:

> **Voice-triggered Condition (Southern Voice, Midland Label) Script:** For the next section, we are going to change the game a bit. The clues will be given to you out loud. Please make sure your headphones are on and volume is at a comfortable level. *With me being from Michigan, hopefully this Midwest accent will be clear and help you guess the words easily*. Our goal with this experiment is to determine whether people are better at guessing words when they are reading alone or interacting with another person. We think interacting with another person will make you better at guessing words, especially if you feel connected to your partner in the game. So here's a little bit of information about me: Name's Jenn, *From Ann Arbor Michigan*, Hobbies include baking and playing guitar. Could you press the record button then briefly state some information about yourself too? Great, thanks! Let's get started with the next section.
>
> **Label-triggered Condition (Midland Voice, Southern Label) Script:** For the next section, we are going to change the game a bit. The clues will be given to you out loud. Please make sure your headphones are on and volume is at a comfortable level. *With me being from Mississippi, hopefully this Southern accent won't cause you any trouble.* Our goal with this experiment is to determine whether people are better at guessing words when they are reading alone or interacting with another person. We think interacting with another person will make you better at guessing words, especially if you feel connected to your partner in the game. So here's a little bit of information about me: Name's Jenn, *From Hurley Mississippi*, Hobbies include baking and playing guitar. Could you press the record button then briefly state some information about yourself too? Great, thanks! Let's get started with the next section.

The addition of these instructions introducing each model talker's dialect is the main way the experiment differed from the design of Experiment 2. While the production data is of primary interest, the lexical decision task measuring perceptual responses was included since participants seemed to show clearer perceptual responses to the experimental manipulations in Experiment 2.

That is, while production data is generally somewhat messier, also including perception data will provide a somewhat clearer picture. The perception component utilized the same lexical decision design with the same Southern model talker heard regardless of exposure condition. Finally, since the goal of this experiment is not to compare individuals but rather to look at whether convergence does or does not occur in each condition, presentation of stimuli was randomized across participants, and the phase in which each word was elicited was also counterbalanced across participants, just as in Experiment 1. Order of presentation of lexical decision stimuli was also randomized across participants.

5.2.3 Analysis

Just as in Experiments 1 and 2, recordings were forced-aligned using the Penn Phonetics Lab Forced Aligner (p2fa) (Yuan and Liberman, 2008). Force alignment was then blindly hand checked and corrected as necessary, with Praat max formant and number of formant settings adjusted as necessary for each participant, and sometimes for each vowel. A Praat script was used to automatically measure the first three formants of target vowels at 7 time points throughout the course of the vowel: 10%, 20%, 30%, 50%, 70%, 80%, and 90%. Formant measurements were then normalized in R using the Lobanov method. Any incorrect responses given were omitted from the analysis, and all data visualization and statistical modeling was done in R (R Core Team, 2015). Just as in Experiments 1 and 2, the measurement of interest we will focus on is the height of the glide (F1 at 80%). Outliers were excluded from the following analysis on a by-item and by-participant basis, with observations greater than 3 standard deviations from the mean being omitted. This leaves a total of 10,018 /aɪ/ tokens (or roughly 87 per person and 29 per-person per experiment phase) in the analysis.

For the perception component, “word” and “non-word” responses on the lexical decision task were collected and analyzed. “Word” responses were coded as 1 and non-word responses as zero. No trials were omitted, and while reaction time data was collected, it will not be analyzed here.

5.3 Results

5.3.1 Production

Two linear mixed effects regression models were fit to the production data, one for each experimental condition, using the LmerTest (Kuznetsova et al., 2017) package in R (R Core Team, 2015), which provides p-values for lmer model fits using Satterthwaite's degrees of freedom method. Separate models seemed to be the best fit for the primary question of interest, which is to determine whether or not convergence occurs within each of the conditions, and whether this differs by dialect background. Further, separate models here ease interpretation of interactions and avoids three-way interactions. The fixed and random effect structure was the same for both models, and was determined by including all model terms that significantly improved fit for each model separately, using likelihood ratio tests. Both models predict normalized F1 at the glide (80%) for each token. Predictors included in the Label-Triggered Condition model are as follows:

- **Phase:** Categorical predictor referring to the experimental phase, a within-subjects manipulation. Levels include Baseline, Exposure (reference level), and Post, treatment coded.

- **Duration:** Continuous predictor referring to token duration, z-scored

- **Frequency:** Continuous predictor referring to lexical frequency, using the log10CD measure from the SUBTLEX corpus, z-scored

- **Dialect:** Categorical predictor referring to a participant's dialect background, with levels Southern (reference) and non-Southern, treatment coded.

- **Phase*Dialect:** Two-way interaction between Phase and Dialect. Because Phase is treatment coded with Baseline as the reference level, and Dialect is treatment coded with Southern as the reference level, the main effect of Phase is restricted only to the Southern participants, and the main effect of Dialect is restricted only to the Exposure Phase.

Random by-participant slopes were also included for Phase and Duration and by-item slopes for Dialect and Duration. The final model for the Label-Triggered Condition data is shown in Table 5.1

Scaled residuals:	Min	1Q	Median	3Q	Max	
	-3.984	-0.604	-0.012	0.552	5.259	
Random effects:						
Groups	Name	Variance	Std.Dev.	Corr		
word	(Intercept)	0.111	0.3332			
	dialectSouth	0.036	0.1892	0.77		
	Duration	0.012	0.111	-0.89	-0.9	
Participant	(Intercept)	0.133	0.3641			
	phaseExp	0.111	0.333	-0.59		
	phasePost	0.243	0.4928	-0.3	0.48	
	Duration	0.044	0.2091	0.27	0.03	0.26
	Residual	0.691	0.831			
Fixed effects:	Estimate	Std. Error	df	t-value	p-value	
(Intercept)	0.074	0.087	101.508	0.846	0.399	
Phase(Baseline)	0.147	0.081	58.515	1.801	0.077	.
Phase(Post)	0.323	0.099	58.599	3.274	0.002	**
Dialect(Non-South)	-0.163	0.093	55.520	-1.752	0.085	.
Duration	-0.401	0.038	105.623	-10.581	<2e-16	***
Frequency	0.107	0.036	69.779	2.948	0.004	**
Phase(Base):Dialect(Non-S)	-0.229	0.108	56.518	-2.115	0.039	*
Phase(Post):Dialect(Non-S)	-0.318	0.131	56.069	-2.434	0.018	*

Table 5.1: Mixed effects linear regression model predicting normalized F1 at 80% for all tokens. Data come from only the Label-Triggered Condition.

For the Voice-Triggered Condition, only duration and frequency significantly improved the model, but the best-fit random effects structure was the same as for the Label-Triggered data. However, since the control variables of duration and frequency were the only fixed predictors significant in the model, and in order to compare the two models directly, Table 5.3.1 below presents the full model for the Voice-Triggered data with the same model specifications as the Label-Triggered data. The results of the models will be discussed in tandem with the corresponding graphs below.

Figure 5.4 plots the mean normalized glide height measurements (F1 at 80%) for all tokens with 95% confidence intervals, broken down by experimental condition and participant dialect background. When we compare participants' production in just the baseline and exposure phases, an interesting pattern emerges. In the Label-Triggered Condition, only Non-Southerners seem to be affected by the experimental manipulation and shift their speech to produce lower, more Southern-shifted /aɪ/ glides after exposure to a talker they have been told is Southern, but is actually from Ohio. Southerners, however, exhibit no such shift in this condition, suggesting that they are not triggered to converge, in the aggregate, through the dialect label alone. The model for the Label-

Scaled residuals:	Min	1Q	Median	3Q	Max	
	-4.332	-0.577	-0.006	0.560	4.584	
Random effects:						
Groups	Name	Variance	Std.Dev.	Corr		
word	(Intercept)	0.120	0.346			
	dialectSouth	0.025	0.159	0.01		
	Duration	0.010	0.102	-0.51	0.34	
participant	(Intercept)	0.089	0.298			
	phaseExp	0.121	0.348	-0.42		
	phasePost	0.090	0.301	-0.34	0.31	
	Duration	0.033	0.183	0.36	-0.33	-0.04
Residual		0.736	0.858			
Fixed effects:	Estimate	Std.Error	df	t-value	p-value	
(Intercept)	0.082	0.078	86.388	1.050	0.297	
Phase(Baseline)	-0.064	0.082	56.817	-0.776	0.441	
Phase(Post)	-0.104	0.076	53.515	-1.372	0.176	
Dialect(Non-South)	-0.179	0.089	55.977	-2.009	0.049	*
Duration	-0.365	0.036	115.987	-10.012	<2e-16	***
Frequency	0.118	0.040	83.922	2.962	0.004	**
Phase(Base):Dialect(Non-S)	0.096	0.108	54.913	0.892	0.377	
Phase(Post):Dialect(Non-S)	0.172	0.101	53.416	1.704	0.094	.

Table 5.2: Mixed effects linear regression model predicting normalized F1 at 80% for all tokens. Data come from only the Voice-Triggered Condition. Only Frequency and Duration significantly improve model fit, but the other predictors included in model 5.1 are included here as well for comparison.

Triggered Condition shown in Table 5.1 confirms that there's no significant shift from Baseline to Exposure for Southerners in this condition (Southerners actually trend toward becoming *less* Southern-shifted after exposure, Est. = 0.147, p = 0.077). There is a significant interaction between Phase and Dialect, suggesting that, compared to Southerners, Non-Southerners produce more Southern-shifted /aɪ/ vowels after exposure (Est.=-0.229, p = 0.039). In the Voice-Triggered Condition, we see the opposite pattern (Figure 5.4), where Non-Southerners do not shift at all during exposure, but Southerners do shift to produce somewhat more Southern-shifted vowels. This shift for Southerners, however, is not statistically significant, as there is no main effect of Phase in the Voice-Triggered Model, nor is there a significant interaction between Phase and Dialect.

Perception

A mixed effects binomial logistic regression model was fit to the perception data, predicting word (1) or non-word (0) responses rates in the lexical decision task. Responses to non-word items are

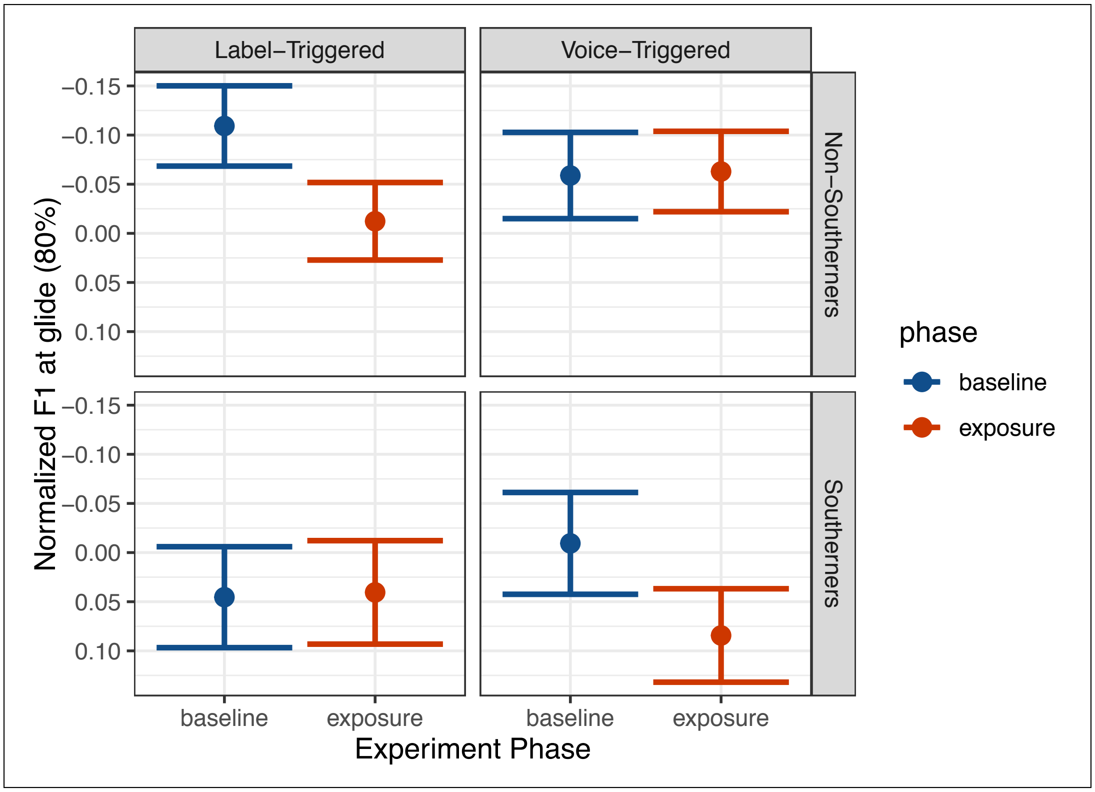

Figure 5.4: Mean normalized F1 at the glide (80%) across all tokens, broken down by participant dialect and experiment phase, with 95% confidence intervals.

not included in the analysis, and only target /aɪ/ items and filler words are compared. Fixed and random effects were included based on likelihood ratio test significance, and the final model is shown in Table 5.3. Fixed predictors included in the model are as follows:

- **Type:** Categorical predictor with levels Target (reference level) and Filler, treatment coded.
- **Condition:** Categorical predictor referring to the voice heard in the exposure phase of the experiment, with levels Voice-Triggered (reference level) and Label-Triggered, treatment coded.
- **Dialect:** Categorical predictor referring to participant dialect background with levels Southern (reference) and Non-Southern, treatment coded.
- **Frequency:** Continuous predictor referring to lexical frequency for each item, using the Log10CD measure from the SUBTLEX corups, z-scored.
- **Type*Condition:** Two-way interaction between Type and Condition. Since both predictors are treatment coded and put into the model as an interaction, the main effect of Condition

Scaled residuals:	Min	1Q	Median	3Q	Max
	-11.878	0.118	0.171	0.25	3.24
Random effects:	Groups	Name	Variance	Std.Dev.	Corr
	participant	(Intercept)	1.791	1.338	
		typefiller	0.96	0.98	-0.91
	word	(Intercept)	0.738	0.859	
		DialectSouth	0.092	0.304	-0.22
Fixed effects:	Estimate	Std. Error	z-value	p-value	
(Intercept)	3.768	0.27	13.975	<2e-16	***
Type(Filler)	-0.122	0.278	-0.44	0.66	
Condition(Label-Triggered)	-0.555	0.258	-2.146	0.032	*
Dialect(Non-S)	-0.816	0.267	-3.056	0.002	**
Frequency	0.298	0.092	3.23	0.001	**
Type(Filler):Condition(Lab-Trig)	0.53	0.23	2.305	0.021	*
Type(Filler):Dialect(Non-S)	0.57	0.243	2.346	0.019	*

Table 5.3: Mixed effects binomial logistic regression model predicting word (1) and non-word (0) response rates on the lexical decision task.

refers to only Target items.

- **Type*Dialect:** Two-way interaction between Type and Dialect. Since both predictors are treatment coded and put into the model as an interaction, the main effect of Dialect refers to only Target items.

The model also includes by-participant intercepts for Type and by-item intercepts for Dialect.

In general, Southerners tend to have higher word endorsement rates for target items than Non-Southerners, and those in the Voice-Triggered Condition tend to have higher word endorsement rates than those in the Label-Triggered Condition. As Figure 5.5 shows, Southerners in the Voice-Triggered Condition (right facet) recognize target items as words at roughly the same rate as filler items, and the lack of significant main effect for Type in the model confirms that there is no significant difference between Target and Filler word endorsement rates for Southerners in this Condition (Est. = -0.122, p = 0.66). The main effect of Condition (Est. = -0.555, p = 0.032) and the significant interaction between Type and Condition (Est. = 0.53, p = 0.021) suggests that those in the Voice-Triggered Condition generally recognize target items at higher rates, and by extension the difference in word endorsement rates between Target and Filler items is smaller in the Voice-Triggered Condition than in the Midland Voice Condition. These effects show that those in the Voice-Triggered Condition show greater shifts than those in the Label-Triggered Condition. There is also a main

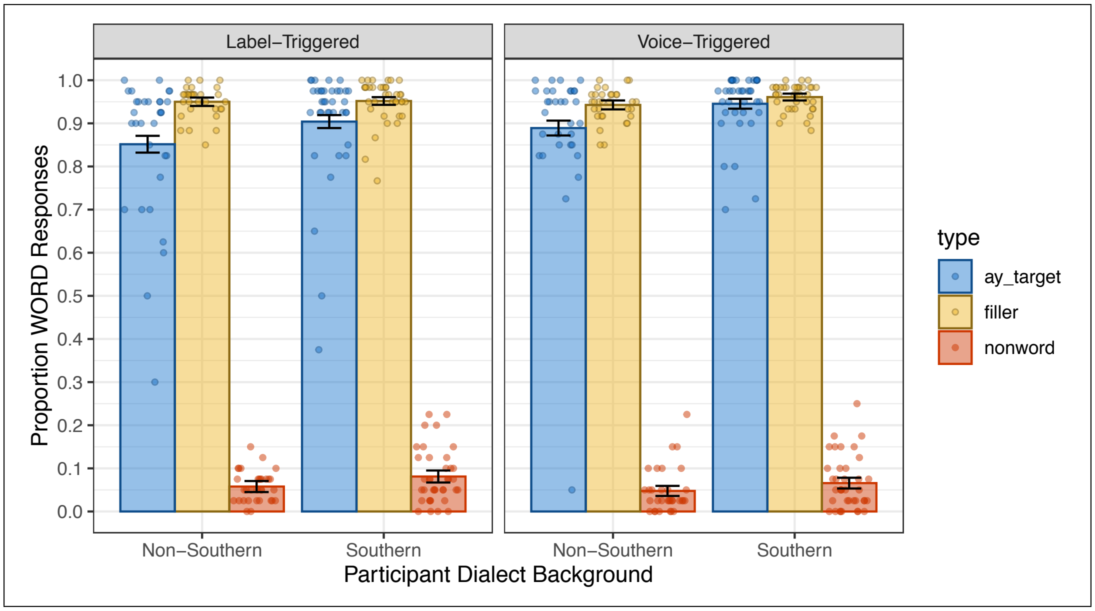

Figure 5.5: Proportion of "word" responses on the lexical decision task across all items with 95% confidence intervals, broken down by Condition, Stimulus Type, and participant Dialect. Points represent participant means, but bar height and error bars are based on all items, not participant means.

effect of Dialect, such that Southerners have higher word endorsement rates for Target items (Est. = -0.816, p = 0.002). Similarly, the significant interaction between Dialect and Type (Est. = 0.57, p = 0.019) suggests that the difference in word endorsement rates between Target and Filler items is smaller for Southerners than for Non-Southerners, again suggesting a greater perceptual shift for Southerners. However, neither a three-way interaction between Type, Condition, and Dialect, nor a two-way interaction between Condition and Dialect significantly improve the model, suggesting that the effect of Dialect is not greater in one condition or the other, and the effect of Condition is not greater for one dialect group.

Southerners' large perceptual response to the Southern-Voice talker aligns with what we see in production, in that Southerners respond to the Southern Voice more than Non-Southerners do in both modalities. While this trend was not significant in production, the effect is much clearer in perception. Also note that, in the Voice-Triggered, Condition Non-Southerners shift less than Southerners. This reinforces the point that Non-Southerners are particularly influenced by the dialect labels. While presence of the Southern dialect label seemed to facilitate Non-Southerners' convergence in production, presence of the Midland dialect label seemed to inhibit convergence in perception, since Non-Southerners exhibit significantly less perceptual shift than Southerners in this

Condition, despite exhibiting perceptual shifts of *the same* magnitude as Southerners in Experiment 2.

What is surprising about the perception data, in light of the production data, is that Non-Southerners do not appear to exhibit any perceptual shift in response to the Southern Label talker (and in fact shift more toward the Southern Voice talker). This may be surprising in light of the fact that Non-Southerners *only* shift toward the Southern Label talker in production. However, it is important to recall that the Southern talker used in the lexical decision task is not the Midland talker that was labeled as "Southern" in the production component. That is, the dialect label of "Southern" was applied to the Midland Voice, which induced convergence in production for Non-Southerners, but when participants moved on to the lexical decision task and heard a Southern speaker (whose dialect was not labeled), they apparently (and unsurprisingly) did not extend the dialect label assigned to the previously heard talker to this new talker. In fact, Non-Southerners in the Label-Triggered Condition have word-endorsement rates around 85%, which is equivalent to those in the Midland-Voice Condition of Experiment 3, suggesting that nothing learned in the exposure phase transferred to the lexical decision task. While hearing a "Southern labeled" Midland talker allowed for generating expectation for *that* talker, no expectations for "Southernness" seem to have been generated for the new talker.

Another somewhat surprising finding is that Southerners *do* show greater perceptual shifts than Non-Southerners in the Label-Triggered Condition, despite exhibiting no shifts in production. This is particularly surprising since, in Experiment 2, Southerners and Non-Southerners in the Midland Voice condition behaved identically on the lexical decision task. I attribute Southerners' increased word endorsement rates in the Label-Triggered Condition to an increased recognition of the lexical decision model talker as Southern after being primed with the notion of "Southernness" in the Word Naming Game. In other words, Southerners in the Label-Triggered Condition would be told they were hearing a Southerner, but would actually be hearing a Midlander, and would not converge in production presumably because there was no bottom-up information supporting this claim; either they didn't believe the talker was really a Southerner, or they didn't think they were Southern-shifted enough to use glide-weakened /aɪ/, or the label alone wasn't enough to induce convergence, so they didn't converge toward this talker. However, once they began the lexical decision phase and hear an actual Southern talker, they may have been primed to more readily interpret this talker as Southern (Compared to Southerners in the Midland Voice Condition of Experiment 1 or 2), and would therefore exhibit perceptual shifts toward the Southern variant. Again, this reinforces the finding that Southerners rely primarily on bottom-up cues. In fact, this interpretation is supported by participants' post-exposure production data, shown in the next section.

5.3.2 Production responses to Lexical Decision Stimuli

The post-exposure data was omitted from Figure 5.4 and the results thus far because it is important to first understand how participants behave in perception in order to understand the post-exposure shifts; participants' post-exposure production seems to be somewhat influenced by the lexical decision task, which immediately preceded the post-exposure phase of the Word Naming Game. Figure 5.6 duplicates Figure 5.4 but adds the data from the post-exposure phase. Consistent with the interpretation that Southerners in the Label-Triggered Condition shift *perceptually* in the lexical decision task because they are responding to the lexical decision talker rather than the Word Naming Game talker, we also see that Southerners shift in *production* after exposure to the lexical decision talker (but not during exposure to the Word Naming Game talker). This reinforces the interpretation that Southerners are driven primarily by bottom-up information, since Southerners here do not shift until they hear the Southern-accented talker (and as mentioned earlier priming the concept of "Southernness" with the exposure talker's dialect label may have facilitated this later influence). The model for the Label-Triggered data, shown in Table 5.1 reveals a significant main effect of Phase for Southerners, such that F1 is higher (more Southern) in the Post-Exposure phase than in the Exposure Phase (Est. = 0.323, p = 0.002). In other words, Southerners shift produce more Southern variants post-exposure. Non-Southerners, on the other hand, remain shifted into the post-exposure phase, but do not shift further. There is a significant interaction between Dialect and Phase for Post-Exposure compared to Exposure (Est. = -0.318, p = 0.0.018, suggesting this shift does not occur to the same extent for Non-Southerners, who have already shifted in the exposure phase.

These post-exposure shifts in the Label-Triggered Condition may seem surprising in light of Experiment 1, which showed that convergence toward expectations lasted only in the exposure phase and participants returned to their baseline post-exposure. However, exposure to actual glide-weakened /aɪ/ tokens in the lexical decision phase may have allowed shifts to last into (or begin in) the post-exposure phase, since as noted in Experiment 2, input-driven convergence triggered by

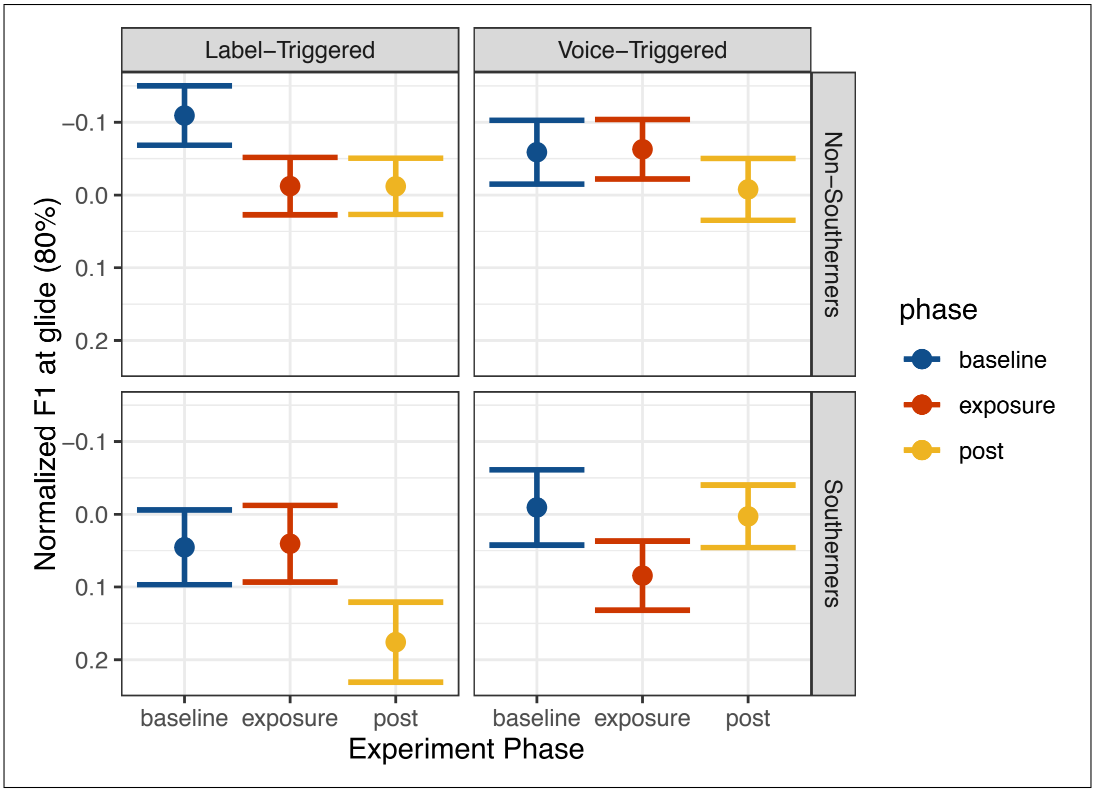

Figure 5.6: Mean normalized F1 at the glide (80%) across all tokens, broken down by participant dialect and experiment phase, with 95% confidence intervals.

actually observing glide-weakened /aɪ/ seems to last longer than convergence triggered by expectations. In the Voice-Triggered Condition, Southerners do not show the same post-exposure shift, perhaps because they were not as influenced by the lexical decision talker since it was the same talker they heard in the exposure phase. Perhaps new input (i.e., observing glide-weakened /aɪ/ in the lexical decision task) does not exert as much influence if expectations about the talker's production for this variant have already been made. Those in the Label-Triggered Condition wouldn't have had this problem because the lexical decision talker was brand new to them, so they had no pre-existing expectations about the talker's /aɪ/ production. This would also explain why the lexical decision talker didn't seem to trigger convergence in the post-exposure phase of the comparable /aɪ/–absent Condition in Experiment 2.

Figure 5.6 also shows that, in the Voice-Triggered Condition (right facets), Non-Southerners shift little across the phases though they produce slightly more Southern variants in the post-exposure phase. This could be due to increased exposure to glide-weakened /aɪ/ words in the lexical decision task, but the shift is minimal and not significant. Southerners, on the other hand, exhibit the expected pattern of expectation-driven convergence we observed in the Southern Voice Condition of Experiment 1: they produce more Southern-like variants during the exposure phase then shift back to their baselines post-exposure. Note, however, that none of the production shifts across phases in the Voice-Triggered Condition are significant for either dialect group.

5.4 Discussion

In summary, Experiment 3 utilized a combined perception-production design that manipulated bottom-up acoustic cues (by using a Midland-accented talker vs. a Southern-accented talker) and contradictory top-down cues (Labeling the Midland talker as "Southern" and the Southern talker as "Midland") to determine the type of cues that trigger expectation-driven perception and production shifts toward Southern speech. In the Label-Triggered Condition, participants were only triggered by the top-down information that the model talker was Southern. In the Voice-Triggered Condition, participants were primarily triggered by the bottom-up acoustic cues from the talker's actual speech. The combined perception-production results suggest that the type of information that triggers convergence differs by dialect group. While Southerners appear to respond primarily to bottom-up cues, Non-Southerners appear to rely primarily on top-down cues. This opposition may reflect the role of linguistic experience in forming sociolinguistic expectations.

5.4.1 Evidence for bottom-up triggers of expectation-driven convergence

The evidence for bottom-up triggers of expectation-driven convergence is most clearly seen in the perception data. Overall, participants who recently heard a Southern-accented talker are more likely to judged glide-weakened /aɪ/ words spoken by the Southern-accented talker as words than when they recently heard a Midland-accented talker, even when the talkers are assigned conflicting dialect labels. This effect is clearest for Southern participants, who recognize /aɪ/ words as words at the same rate as filler words if they've previously been exposed to the Southern-accented talker. Further, in production, only Southern participants shifted to produce more Southern-like variants of /aɪ/ after hearing a Southern voice, though this shift was not statistically significant. It is potentially possible for participants to have extrapolated the Southern-accented talker's expected /aɪ/ glide values based on the position of other vowels in the vowel space (as suggested in Section 5.1.2), though that does not appear to be the best explanation here due to the fact that Southerners are the ones primarily influenced by bottom-up cues. Both dialect groups should be able to fill in predicted vowel values based on the position of other vowels in the vowel space at equally high rates. However, since Southerners seemed to be influenced by bottom-up cues more than Non-Southerners here, this suggests that their advantage stems from previous experience with the Southern dialect and greater exposure to statistical co-occurrence patterns among glide-weakened /aɪ/ and other Southern shifted vowels.

Finally, even the Southerners in the Label-Triggered Condition eventually shifted to produce more Southern-like variants, but crucially only in the post-exposure phase after hearing the lexical decision task model talker who was Southern-accented and *did* produce glide-weakened variants of /aɪ/. This exposure to the variant of interest in the lexical decision task likely facilitated convergence in the post-exposure phase, but it's also possible that convergence was facilitated by priming Southern participants with the concept of "Southernness" via the dialect label. While the Southern dialect label alone was not enough to cause Southerners to shift, the dialect label coupled with the later exposure to Southern-accent talker seemed to be sufficient for inducing shifts in both perception and production.

From these results alone, it's unclear whether bottom-up shifts occur without *some* activation of the concept of "Southernness," since participants may be positing their own dialect labels based on acoustic cues. Put another way, it's unclear whether bottom-up information alone is enough to trigger expectations, or if it's even possible to use bottom-up acoustic cues associated with Southernness without positing the social category of "Southern." Regardless, Southerners do not seem to exhibit shifts based on dialect labels *alone*, and we can conclude that bottom up cues are able to trigger expectation-shifts.

5.4.2 Evidence for top-down triggers of expectation-driven convergence

Experiment 3 has also provided evidence that top-down cues alone, in this case dialect category labels, can induce convergence, though this only seems to be the case for Non-Southerners. This finding is most apparent in the production data. In the Label-Triggered Condition, Non-Southern participants shifted to produce more Southern-like variants of /aɪ/ during exposure to a model talker whom they were told was from the South, but who actually had a Midland accent. Non-Southerners did not show similar shifts when they *did* actually hear a Southern-accented talker, but were told the talker was from the Midwest, suggesting heavy reliance on dialect labels rather than bottom up acoustic cues.

Non-Southerners, however, do not show corresponding shifts based on dialect labels in perception, though this is likely because the model talker they hear for the perception task is actually the Southern-shifted model talker, and not the Midland-accented talker they had previously been listening to, who had a "Southern Dialect" label. Failure to generalize perceptual learning to a novel voice is consistent with findings in the perceptual learning literature (Eisner and McQueen, 2005; Kraljic and Samuel, 2005). Though note that some studies have found learning to generalize to novel talkers if they are similar enough, such as of the same gender (Reinisch and Holt, 2014). It's likely that, though both model talkers were female, their voices and dialects were dissimilar enough to prevent generalization from one to the other. This also brings up another important point, which is that activating the concept of "Southernness" in and of itself does not seem to be sufficient for eliciting convergence (though findings that social concepts can trigger shifts in production or perception have previously been reported (Hay and Bresnan, 2006; Hay and Drager, 2010; Sanchez et al., 2015; Love and Walker, 2013)). Rather, individuals seem to need to integrate this social concept into their representation of their interlocutor in order to converge. Note, however, that it's possible that such "social-concept priming" effects may have contributed to Southerners being more likely to shift perceptually to a novel talker after being recently exposed to the concept of Southernness, as discussed above, but this priming only made them more likely to notice the dialect of the Southern model talker.

Finally, just as presence of the Southern dialect label can facilitate shifts, a contradictory label seems to inhibit shifts (though only for Non-Southerners who rely on this top-down information; the Midland label does not seem to inhibit Southerners' perceptual shifts, perhaps because they do not rely primarily on top-down information to begin with). This may explain why there was no dialect effect in perception in Experiment 2—both dialect groups shifted equally and Non-Southerners seemed to shift perceptually based on the voice of the Southern talker alone—yet Southerners shift more toward the Southern voice here. In Experiment 2, Non-Southerners were free to posit their own dialect labels based on acoustic cues, while here in Experiment 3, the Midland label seems to particularly inhibit Non-Southerners, such that they shift less perceptually than Non-Southerners in

Experiment 2 who had no dialect label at all.

5.4.3 The role of dialect experience

Both top-down and bottom-up information seem to be relevant for inducing shifts toward Southern speech in perception and production, though it is interesting that the two dialect groups tend to rely on these cues differently. This begs the question, why did Southerners not shift where Non-Southerners did, and vice versa? Different types of experience with the Southern dialect likely contribute to these differences.

It is interesting that Non-Southerners did not exhibit shifts based on bottom-up cues signalling "Southernness" when dialect labels contradicted these cues. We might speculate that this has to do with the relatively greater social salience of glide-weakened /aɪ/ compared to other Southern-shifted features. Previous research on recognition of the Southern accent has shown that glide-weakened /aɪ/ is a particularly salient feature indexing Southernness, and other Southern-shifted features are not as well recognized (Labov, 2010; Torbert, 2010). Non-Southerners, who have less experience with Southern speech, may not associate non-/aɪ/ features observed from the Southern-accented talker with "Southernness," and therefore may not have formed expectations for glide-weakened /aɪ/ from this talker, particularly since some other features of Southern-shifted speech occur in other dialects of English, such as back vowel fronting. This may also explain why Non-Southerners had weaker production shifts than Southerners toward the unlabeled Southern-accented talker in Experiment 1. However, when participants are explicitly prompted with the label of "Southernness," if /aɪ/-glide weakening is one of the most salient linguistic features associated with this social category, this would trigger convergence toward glide-weakened /aɪ/ better than exposure to other Southern vowels would.

On the other hand, Southerners' failure to shift based on dialect labels alone may reflect a greater degree of experience with Southern speech, which may contribute to greater awareness that just because a talker is labeled as "Southern" does not mean they will use glide-weakened /aɪ/, and that a more accurate way to form expectations is through observation of how Southern-shifted the talkers other vowels are. In other words, Southerners may be more aware that being from the South does necessitate Southern-accentedness (after all, most of these Southern participants were not Southern-shifted themselves). A similar (more automatic account that does require awareness for bottom-up cues to trigger shifts in perception or production) is that the concept of "Southern" did not as strongly activate "glide-weakened /aɪ/" variants for Southerners because they would presumably have experienced many instances of Southerners *not* using glide-weakened /aɪ/, including in their own speech. Further, the concept of "Southernness" may have more nuanced associations for Southerners, and may thus not be as strongly linked to this one particular linguistic variant. Non-Southerners, in contrast, may only be aware that a talker is Southern if cued by their accent, and would therefore have more associative links between "Southerner" and "glide-weakened /aɪ/." Of course, another possibility is that many Southerners simply did not believe the "Southern" dialect labels and failed to shift for this reason.

Chapter 6

General Discussion and Conclusions

The dissertation investigated language users' perceptual and production responses to both observed and expected linguistic behavior using a series of experiments. Production responses were measured via participants' linguistic convergence toward (or divergence from) a model talker, using a novel Word Naming Game paradigm developed for the dissertation; perceptual responses were measured with a perceptual learning paradigm using an auditory lexical decision task designed to elicit perceptual category shifts to encompass an observed or expected variant. The variant of interest across all three experiments was glide-weakened /aɪ/, a salient feature of Southern U.S. English. Experiment 1 (Chapter 3) found that individuals exhibit expectation-driven convergence and produce more glide-weakened /aɪ/, which they may reasonably *expect* but never directly observe from a Southern-shifted model talker. Experiment 2 (Chapter 4) compared input-driven and expectation-driven shifts in both perception and production, comparing individual behavior across perception and production tasks when participants only *expect* vs. actually *observe* glide-weakened /aɪ/ from a model talker. This experiment found that input-driven and expectation-driven shifts in both perception and production pattern somewhat differently, and while no clear perception-production correlation was observed on an individual level, individual patterns hint toward the possibility that perceptual responses are a necessary, but not sufficient, precursor to convergence. Finally, Experiment 3 (Chapter 5) used a dialect-label manipulation to investigate whether top-down social information (a non-Southern talker labeled as "Southern") or bottom-up acoustic cues (a Southern-shifted talker labeled as "Non-Southern") trigger expectation-driven shifts. This experiment found that Non-Southerners are cued by top-down information, while Southerners rely on bottom-up acoustic cues. Taken together, the experiments presented in this dissertation make several important empirical, methodological, and theoretical contributions to research on linguistic convergence, perception-production relationships, and the nature of sociolinguistic knowledge.

For one, the dissertation has drawn a theoretical and terminological distinction between two types of convergence that vary in the relationship between the trigger of convergence and the linguistic feature targeted for convergence. The type of convergence previously observed in experimental work on convergence has been specified here as *input-driven* convergence, which differs from (previously only anecdotally observed) *expectation-driven* convergence, during which individuals do not directly observe (but only expect) the linguistic variants they converge toward. In addition to this terminological distinction, the dissertation has established expectation-driven convergence as a real, empirically-observable phenomenon that has replicated under somewhat different circumstances. This finding supports anecdotal accounts of the existence of expectation-driven convergence (Bell, 1984), and also aligns nicely with the findings of Wade and Roberts (2020), who found evidence of expectation-driven convergence in computer-mediated communication in an artificial "alien" language. The dissertation also contributes to our understanding of the nature of such convergence and its relationship to input-driven convergence.

In order to observe expectation-driven convergence in a controlled way, I devised new experimental methodologies, including the Word Naming Game task for eliciting targeted speech data. Further, a lexical decision task used to investigate perceptual learning behavior was integrated into the Word Naming Game task to investigate expectation-driven shifts in production and perception simultaneously. These methods have proved successful in eliciting shifts in both perception and production toward expected and observed linguistic variants, and may prove to be useful for eliciting speech and assessing perceptual shifts to answer other empirical questions both within and beyond the realm of linguistic convergence.

Finally, in establishing the existence of expectation-driven convergence and some of its properties, the dissertation has made contributions to our understanding of (1) the mechanisms underlying linguistic convergence (2) the relationship between linguistic perception and production and (3) the nature of social and linguistic representations, and their mental links. These theoretical contributions are discussed in greater depth in the following subsections.

6.1 The mechanisms underlying linguistic convergence

By directly comparing input-driven and expectation-driven convergence, isolating the top-down and bottom-up triggers of expectation-driven convergence, and examining the perceptual underpinnings of convergence, the experiments presented here allow us to draw some conclusions about the mechanisms underlying linguistic convergence, as well as the relationship between linguistic perception and production.

Firstly, the existence of expectation-driven convergence suggests that *not all* instances of linguistic convergence can be straightforwardly explained via automatic perception-production relationships where production is directly derived from perception (e.g., Motor Theory, Liberman and Mattingly (1985), Direct Realism, Fowler (1986), some episodic accounts of convergence such as Goldinger (1998)). As illustrated in the dissertation, convergence may occur toward a variant not directly observed in the input, and accounts of convergence relying on tight perception-production feedback loops or other real-time perception of the target variant are not well-suited to explain this behavior. This is not to say that convergence is *never* accomplished via these automatic perception-production relationships. In fact, one possibility is that input-driven convergence and expectation-driven convergence are governed by distinct cognitive mechanisms. Experiment 2 provided some evidence that input-driven and expectation-driven shifts do pattern differently: input-driven productions shifts seem to last longer and some individual difference measures such as Social Desirability or Openness scores have opposite predictions for input-driven and expectation-driven convergence. Further, input-driven perceptual shifts were universal across participants, while some participants did not exhibit expectation-driven perceptual shifts; this may be because input-driven convergence stems from automatic perception-production linkages while expectation-driven convergence requires some amount of awareness and control. Although, of course, another explanation for the variable shifts in perception in the expectation-driven (/aɪ/–absent) condition is that those who failed to shift perceptually simply did not have the necessary experience to form expectations based on non-/aɪ/ triggers.

A second possibility is that expectation-driven and input-driven convergence are governed by the same cognitive mechanisms, but these mechanisms are simply triggered by different types of information. Exemplar Theory, and other connectionist models of linguistic processing that allow for encoding of social and contextual information alongside detailed phonetic information, are well-suited to explain how sociolinguistic associations are formed and can explain how phonetically-detailed representations of glide-weakened /aɪ/ may be activated via various associative links, whether linguistic (such as other Southern-shifted vowels frequently encountered alongside glide-weakened /aɪ/, or glide-weakened /aɪ/ itself during input-driven convergence) or social (such as the label of "Southern" frequently encountered alongside glide-weakened /aɪ/). Usage-based models propose that production targets are derived from (activated) exemplar clouds, predicting that production shifts would automatically fall out of recent activation of glide-weakened /aɪ/ (Pierrehumbert, 2001). It is possible, then, that both expectation-driven and input-driven shifts in production are automatic consequences of activation of glide-weakened /aɪ/, though this activation may be triggered by different cues. This would also account for the possibility that top-down social cues or bottom-acoustic cues could trigger expectation-driven convergence to glide-weakened /aɪ/. However, we can also imagine several ways in which production may not align directly with perception. Indeed, convergence and divergence behaviors seem to stem from similar perceptual shifts, suggesting convergence is not an automatic consequence of perception, and models of linguistic convergence must be able to account for such misalignment.

It's likely that any automatic influences on production may be inhibited or otherwise modulated by social factors such as desire for affiliation or to maintain social distance. That is, while activation of /aɪ/ may be an automatic consequence of direct perception of associated cues, resulting production shifts *may* be automatic or may require some degree of awareness or control. This could be part of the reason we seldom see universal shifts in production in studies on linguistic convergence, even when the variant of interest is directly observed. There are a number of ways that automaticity and control may interact in perception and production, which are future directions worth exploring. While we cannot say with certainty whether input-driven convergence and expectation-driven convergence are governed by the same cognitive mechanisms, models that simultaneously account for both behaviors are appealing when explaining linguistic convergence in that they can account for the multiple triggers of convergence observed in the dissertation (Southern dialect labels, other Southern-shifted vowels, or glide-weakened /aɪ/ itself), as well as other related phenomena showing that social factors influence linguistic processing and vice versa.

Another reason that shifts in perception may be inhibited in production is an attempt to preserve contrasts between phonemic categories. Experiment 2 revealed substantial divergence among Southern participants, which was attributed to the fact that many more Southern participants in Experiment 2 were already Southern shifted, compared with the other two experiments. In both Experiments 1 and 2, participants' baseline productions mediated convergence behavior, such that already somewhat Southern-shifted participants converged less, or even diverged, from the Southern model talker. It does not seem to be the case that Southern-shifted participants simply failed to *notice* the variant because it was not sufficiently different from their own speech; on the contrary, Southerners (and divergers) shifted just as much in perception as other participants. The most likely explanation is that participants with somewhat glide-weakened /aɪ/ to begin with could not shift much further toward the model talker without threatening contrast preservation between /aɪ/ and /ɑ/ or /æ/ categories. Active inhibition of convergence may have ultimately resulted in divergence due to a desire to preserve phonemic contrasts. All in all, these findings suggest that production shifts cannot be entirely an automatic consequence of perception, and inhibition or facilitation of convergence may stem from both internal and external factors.

6.2 Implications for a mental model of sociolinguistic cognition

The results of the dissertation provide some empirical observations that should be accounted for in our understanding of how social and linguistic knowledge are mentally represented and interact. For one, the dissertation reinforces previous findings that both top-down social information (e.g., D'Onofrio, 2015; Hay and Drager, 2010; Niedzielski, 1999) and bottom-up acoustic cues (e.g., Theodore and Miller, 2010) can influence processing of associated phonetic variants. The results of Experiment 3 specifically show that macro-level social categories such as "Southern" as well as acoustic details (other Southern-shifted vowels) are also stored in memory and linked to associated phonetic variants. These findings suggest strong links between both top-down and bottom-up information and phonetically-detailed linguistic representations that can be accessed rapidly and accurately. Additionally, since such associative links have been shown across all three experiments (and in previous literature) to influence a range of different linguistic processes such as speech perception, speech production, and expectation-formation, a model of sociolinguistic cognition must also allow for these different processes to similarly utilize sociolinguistic representations.

Further, the results of the dissertation support a view in which both automatic as well as more controlled and intentional linguistic behaviors are similarly influenced by associative links between social and linguistic representations, and as such our model of sociolinguistic cognition should be able to account for both types of responses. The experiments presented in the dissertation demonstrated that social knowledge guides linguistic expectation formation and influences perceptual responses (measured by lexical judgments), which were almost universal, likely indicating rapid and automatic integration of social knowledge. The influence of activated social knowledge on linguistic perception and production is easily accounted for with the sort of rapid, automatic links proposed in exemplar models or social priming accounts. But, as Campbell-Kibler (2016) notes, "The linkage of social information to grammatical structures and stored linguistic exemplars, however, does not obviate our need for a resource-heavy, attention-based process along the lines of the sociolinguistic monitor." That is, automatic influences are only one piece of the puzzle, since "some sociolinguistic shifts occur apparently independently of conscious introspection, but other sociolinguistic behavior appears effortful, poorly integrated with other linguistic structure and available to verbally accessible control" (p. 141). Indeed, the dissertation also suggests that socially-triggered expectations interact with both social motivations (measured in terms of an individual's social desirability) as well as structural pressures (such as divergence shifts as a means of phonemic contrast preservation) in a way that presumably involves more attention, awareness, and control. A holistic model of sociolinguistic cognition must then be able to account for both automatic and more controlled influences of social knowledge on linguistic behavior.

The dissertation also made several observations suggesting that activation of general social concepts may not influence linguistic behavior in the same was as social characteristics understood *as being about a talker*, and this observation may also have implications for models of sociolinguistic cognition. For instance, there is some evidence of interlocutor specificity in perception and production behaviors. For one, expectation-driven convergence in Experiment 1 was confined to the exposure phase alone, and participants promptly returned back to their baseline productions when they discontinued exposure to the model talker. However, some longer-lasting convergence effects were observed for input-driven convergence. Secondly, Non-Southerners' production shifts toward the model talker labeled as "Southern" did not persist when the model talker switched to an actual Southerner who was not given a Southern label. This suggests that simply evoking the *concept* of

Southernness does not yield associated perceptual shifts unless that label is *applied* to a particular talker. While this contrasts with some sociolinguistic research showing that social concepts alone are enough to induce shifts in linguistic perception (e.g., Hay and Drager, 2010; Sanchez et al., 2015), this may be because social concepts applied to an interlocutor induce *stronger* shifts than social concepts alone, which may not have elicited large enough responses to be detected in this study. Our models of sociolinguistic cognition should be able to account for evidence that information that is understood as being about a person is structured differently and retained better than information that is not attributed to a speaker (e.g., Asch, 1946; Chartrand and Bargh, 1996).

Finally, the results of Experiment 3 in particular also provide insights into how linguistic experience may shape sociolinguistic representations, which is important to consider in modeling how sociolinguistic representations are acquired in the first place. For one, it seems to be the case that associative patterns among co-occurring phonetic variants require substantial exposure in order to be acquired—as only those who have spent most of their lives living in the South seemed to activate glide-weakened /aɪ/ after exposure to other Southern-shifted forms alone. Further, it is likely that more salient linguistic forms (such as glide-weakened /aɪ/) may have stronger associative ties to social categories than less salient linguistic forms or linguistic forms that are also linked to other social categories (such as fronted back vowels, which may be associated with other dialects of English). This may be why Non-Southerners responded more to the Southern label than to the Southern voice that didn't contain any /aɪ/ tokens. Finally, regarding the nature of associative links between social and linguistic information, we may consider the possibility that mental links between labels and linguistic forms can be washed out or attenuated by counter-evidence. For instance, "Southern" may indicate /aɪ/-glide weakening for Non-Southerners, who are potentially only aware of an individual's "Southern" designation based on their Southern-shifted dialect features; whereas Southerners, who are around people they recognize as "Southern" every day, may see more counter-evidence that not all Southerners are Southern-shifted, so "Southern" may not link automatically to glide-weakened /aɪ/ since these links may be weaker and the concept of "Southernness" may be more strongly tied with other associations.

6.3 Future Directions

The dissertation contributes to a growing body of research in sociolinguistic cognition showing that social and linguistic knowledge are inextricably linked in the minds of speakers. The dissertation is just a small step toward a better understanding of this relationship, but it lays some groundwork for answering new questions about sociolinguistic cognition and highlights a number of promising areas for future research.

For one, the methods developed and utilized here for investigating expectation-driven shifts in perception and production could be adapted and used diagnostically to understand what types of expectations individuals posit based on various social or linguistic cues. Eliciting individuals' sociolinguistic expectations has been accomplished using a number of experimental paradigms, including eye-tracking ((Koops et al., 2008)), the Matched Guise Technique (Campbell-Kibler, 2011, 2009) and vowel identification tasks (Niedzielski, 1999; Hay and Bresnan, 2006) to name a few. Of course, each method brings with it its own difficulties and task-demands, so adding to this arsenal of sociolinguistic methods will be useful for testing reliability in eliciting the same linguistic behavior across experimental paradigms. Further, testing the reliability of expectation-driven convergence toward features at different levels of linguistic structure (perhaps those not so salient) and in different language varieties may help to shed light on the role of other factors such as salience in expectation formation and will certainly broaden our understanding of these novel phenomena, as well as linguistic convergence generally.

The dissertation also brings up a number of unanswered questions, and speculates about explanations for some observed behaviors that can be empirically tested in future work. For instance, the perceptual correlates of convergence still remain somewhat unclear. Descriptive analysis of individual perception-production patterns hints that perceptual shifts may be necessary for both convergence and divergence (but not sufficient for eliciting either), which aligns with Kraljic et al's (2008) interpretation of a lack of obvious correlation between individuals' perception and production shifts toward the same variant. A wider range of studies on the perceptual correlates of convergence—particularly utilizing different a wider range of perception and production tasks—will help to determine the generalizability of these interpretations.

Another unanswered question concerns the relative automaticity/control of convergence and the perception-production relationship in general. While this has been a long-standing question in research on linguistic convergence without a clear answer, the dissertation has made some steps toward better understanding this relationship. I speculated that divergence may stem from the same perceptual shifts as convergence based on evidence that all convergers and divergers (but not maintainers) shifted in perception, and some individual differences measures similarly predicted convergence and divergence behaviors, which again differed from maintainers' behaviors. I suggest here, that by giving more attention to divergence and maintenance behaviors, rather than disregarding them as simply divergent from aggregate group norms in studies on linguistic convergence, we may gain a better understanding of what aspects of linguistic behavior are automatic or require some degree of awareness or control.

While the dissertation found that several of the same individual differences predictors previously shown to be relevant for predicting linguistic convergence were also relevant for the present study, the exploratory investigation into individual differences predictors of expectation-driven and input-driven convergence yielded some surprising results. For instance, Social Desirability and Openness oppositely predicted input-driven and expectation-driven convergence. Future work on the role of individual differences in convergence may benefit from more work on how different individual measures may contribute to different components of convergence.

To conclude, this dissertation serves as a step towards a better understanding of the mental representations of both linguistic and social knowledge—as well as the ways in which these representations are linked. The findings presented throughout the dissertation contribute to a growing body of research within sociolinguistics, and sociolinguistic cognition specifically, which promotes the understanding that social and linguistic knowledge are inextricably linked. This conclusion highlights the importance of more research into the nature of sociolinguistic expectations and the mental relationship between social and linguistic knowledge, not only to gain a more holistic understanding of linguistic cognition, but also because such work may have broader social impacts. Demonstrating the mental links between social and linguistic knowledge reinforces the social impacts of behaviors such as linguistic discrimination and linguistic evaluation, and future work on sociolinguistic expectation formation specifically may contribute to a better understanding of how social stereotypes are formed, mentally represented, and utilized. Finally, it is my hope that this line of work will not only seek to tackle the many questions that the dissertation has left unanswered, but that it may also contribute to a greater focus on the role of social factors and variability in linguistic cognition and the importance of cognition within sociolinguistic theory.

Appendices

Appendix A: Follow-up Survey Questions

Please rate the following statements about SOUTHERNERS and/or SOUTHERN speech.

Strongly Disagree	Disagree	Somewhat Disagree	Neither agree nor disagree	Somewhat Agree	Agree	Strongly Agree

0 10 20 30 40 50 60 70 80 90 100

I have lived in the South

I am familiar with the way Southerners speak

I have many friends or family from the South

I would like to live in the South

I like the way Southerners talk

I have a Southern accent

My parent(s) have a Southern accent

The majority of people in my school/job have/had a Southern accent

I can imitate a Southern accent well

Please rate the following statements about the person you heard **giving clues** during the experiment

Strongly Disagree		Disagree	Somewhat Disagree		Neither agree nor disagree		Somewhat Agree	Agree		Strongly Agree
0	10	20	30	40	50	60	70	80	90	100

The talker sounded like they are from the South

The talker sounded like they are from the Midwest

The talker sounded kind

The talker sounded friendly

The talker sounded intelligent

The talker sounded wealthy

The talker sounded professional

The talker sounded attractive

The talker is someone I would like to be friends with

The talker sounded happy

The talker is someone I could relate to

The talker sounded educated

Appendix B: Items Elicited in Word Naming Game Task

Elicited Words

	List A	freq	syllables	List B	freq	syllables	List C	freq	syllables	fol_seg
1	my	3.9217	1	I	3.9229	1	why	3.9191	1	#
2	hi	3.7777	1	guy	3.8192	1	try	3.8484	1	#
3	die	3.6682	1	buy	3.6322	1	lie	3.5256	1	#
4	cry	3.279	1	dry	3.187	1	fly	3.3259	1	#
5	sky	3.0913	1	tie	3.1697	1	pie	2.8982	1	#
6	shy	2.842	1	spy	2.7126	1	fry	2.5185	1	#
7	rye	2.1492	1	thigh	2.1847	1	sly	2.0453	1	#
8	bye	3.5408	1	high	3.6485	1	eye	3.5075	1	#
9	bribe	2.382	1	tribe	2.2856	1	fiber	2.0755	2	b
10	ride	3.4937	1	side	3.6745	1	hide	3.3709	1	d
11	wide	2.9926	1	bride	2.8055	1	pride	2.9717	1	d
12	cider	1.699	2	idaho	1.3617	3	idol	2.0128	2	d
13	Friday	3.1245	2	spider	2.365	2	hydrogen	1.9031	3	d
14	slide	2.7559	1	tide	2.4472	1	divide	2.3945	2	d
15	lime	2.0934	1	rhyme	2.2068	1	dime	2.6758	1	m
16	climb	2.80	1	crime	3.279	1	prime	2.738	1	m
17	primate	1.4472	2	chime	1.5911	1	mime	1.5441	1	m
18	find	3.8858	1	kind	3.8644	1	mind	3.837	1	n
19	line	3.6571	1	mine	3.7346	1	fine	3.8535	1	n
20	nine	3.2945	1	wine	3.1816	1	sign	3.5465	1	n
21	dinosaur	2.0682	3	china	2.8007	2	minus	2.3997	2	n
22	vinyl	1.716	2	rhino	1.6128	2	sinus	1.8573	2	n
23	spine	2.3838	1	vine	1.9085	1	pine	2.3263	1	n
24	pint	2.0899	1	grind	2.2227	1	hind	1.7559	1	n
25	benign	1.699	2	canine	1.8573	2	feline	1.5441	2	n
26	behind	3.6749	2	blind	3.1316	1	time	3.9195	1	nasals
27	five	3.7165	1	alive	3.5724	2	drive	3.57	1	v
28	chives	1.1761	1	hive	1.5441	1	dive	2.5198	1	v
29	ivy	1.9191	2	saliva	1.9685	3	ivory	1.8062	3	v
30	rise	3.0039	1	wise	3.0199	1	size	3.2294	1	z
	mean	2.7781	1.34615384		2.7570766666	1.3		2.78133666666	1.33333333	
	std dev	0.8179476309			0.7962402031			0.79108212986		

word	vowel	hint	clue
alive	ay	a _ _ _ _	This word means the opposite of 'dead.'
behind	ay	b _ _ _ _ d	This is the opposite of 'in front of.'
benign	ay	b _ _ _ _ n	This word can be used in reference to a tumor that is not malignant.
blind	ay	b _ _ _ _	The word 'deaf' is used when someone cannot hear. This word is used when someone cannot see.
bribe	ay	b r _ _ _	This word refers to a sum of money or gift given to a person to persuade them to behave a certain way.
bride	ay	b _ _ _ _	This is the term for a woman on her wedding day. She may wear a gown and veil and marry the groom.
buy	ay	b _ _	This word means the same as 'purchase.'
bye	ay	b _ _	This is the opposite of 'hello.'
canine	ay	c _ n _ _ _	This is a term for a dog or related animals.
chime	ay	c h _ _ _	This word refers to a bell or metal tube that makes melodious ringing sounds when struck or blown in the wind.
china	ay	c _ _ _ _	This country is part of Asia and home to urban centers such as Beijing.
chive	ay	c h _ _ _	This is another word for 'scallion' or 'green onion.'
cider	ay	c _ _ _ r	This word is a beverage that can be alcoholic. It is usually made of apples.
climb	ay	c _ _ _ _	This is the action someone performs when going up a ladder or a mountain.
crime	ay	c _ _ _ _	This word refers to any unlawful act. Someone who has robbed a store or committed arson, for example, can be said to have committed one of these.
cry	ay	c _ _	This word is what people do when they are sad. It may involve tears and sobbing.
die	ay	d _ _	This word means to stop living.
dime	ay	d _ _ _	This US coin is small, silver, and worth ten cents.
dinosaur	ay	d _ _ _ _ _ _ _	This term is used for a group of prehistoric creatures, including T-rex, Stegosaurus, and Velociraptor.
dive	ay	d _ _ _	This word refers to the action of plunging head first into water, often from a board at the edge of a swimming pool.
divide	ay	d _ v _ _ _	Similar to 'add' or 'subtract,' this term refers to splitting into different parts.
drive	ay	d _ _ _ _	To get yourself from one place to another in your car, you do this action.
dry	ay	d _ _	This word may be used to mean the opposite of 'wet.'
eye	ay	e _ _	This word refers to the body part that allows us to see. It contains a pupil and cornea and has lids and lashes.
feline	ay	f _ _ _ n _	This word refers to animals in the cat family.
fiber	ay	f _ b _ r	This word refers to a nutrient that may be soluble or insoluble. Fruits and bran are rich in this nutrient.
find	ay	f _ n _	If you've lost something, you probably hope that you will be able to _____ it eventually.
fine	ay	f _ _ _	You may respond with this word when someone asks how you are doing. You may also have to pay one of these if your rented books or videos are overdue.
five	ay	f _ _ _	This number comes between 4 and 6.
fly	ay	f _ _	Birds and planes both perform this action to travel through the air.
friday	ay	f _ _ _ _ _	This is the name for the day of the week that comes before Saturday and for many marks the end of the work week.
fry	ay	f _ _	This refers to a way of cooking something in oil. Foods cooked this way usually turn out greasy.
grind	ay	g r _ _ _	This is what you would do to your coffee before brewing it if it came as whole beans rather than pre-ground beans.
guy	ay	g _ _	This word is used to refer to a male person. It may be used interchangeably with 'man', 'boy', 'dude', or 'fellow.'
hi	ay	h _	This word may be used as a greeting. A shorter way of saying hello.
hide	ay	h _ _ _	This word is part of the name of a well-known game that children usually play: ____ -and-Seek.
high	ay	h _ _ _	This word means the opposite of 'low.'
hind	ay	h _ _ _	You may use this word to refer to a body part in the rear of the body. For example, a dog has two front legs and also ____ legs, meaning they are in the back.
hive	ay	h _ _ _	This word refers to the place where bees live.
hydrogen	ay	h _ _ _ _ _ _ n	This is the element the H stands for in the chemical compound H2O, or water.
I	ay	_	This word is a single letter pronoun that is used to refer to oneself.
idaho	ay	i _ _ _ _	This is a US State known for producing potatoes.
idol	ay	i _ _ _	American ______ is a popular TV show where singers compete in front of judges. Kelly Clarkson is an early winner of this show.

word	vowel	hint	clue
ivory	ay	i _ _ r _	A hard cream-colored material from the tusks of animals such as elephants. Hunting animals for this valuable substance and selling it is illegal.
ivy	ay	i _ _	This is a green, leafy plant that often scales up buildings. Also part of the name of a subset of prestigious universities, the _____ league.
kind	ay	k _ _ _	This adjective may be used to suggest that a person is good-natured or sweet.
lie	ay	L _ _	This word has the same meaning as 'fib' and also as 'untruth'.
lime	ay	L _ _ _	This is a green citrus fruit similar to a lemon. It's the traditional flavor of margaritas.
line	ay	L _ _ _	The shortest distance between two points is a straight _______ .
mime	ay	m _ _ _	This word refers to a performer who uses gestures to communicate but does not speak. Often, they wear gloves and face paint.
mind	ay	m _ _ _	This word refers to the part of a human that thinks, feels, perceives, and reasons. Often used as another word for 'brain.'
mine	ay	m _ _ _	When you own something, you would say that it is 'yours.' When it belongs to me, it is this word.
minus	ay	m _ _ _ _	This is a mathematical notation. The plus symbol is used for addition, and this symbol is used for subtraction.
my	ay	m _	This word may be used to show possession. The word 'your' shows it belongs to you. This word shows it belongs to me.
nine	ay	n _ _ _	This number comes between 8 and 10.
pie	ay	p _ _	This dessert is similar to a cake. It has a crust and comes in many different flavors such as cherry, apple, or banana creme.
pine	ay	p _ _ e	This word may refer to a genus of tree. It usually has needles.
pint	ay	p _ _ t	Sixteen fluid ounces is equivalent to this unit of measurement. Also a way of ordering a large beer in a glass.
pride	ay	p r _ _ _	This word is part of the name of a well-known Jane Austen novel: _____ and Prejudice.
primate	ay	p r _ m _ _ _	This word refers to a group of mammals including monkeys, apes, and lemurs.
prime	ay	p _ _ _ _	Amazon's subscription service that offers free 2-day delivery and other services.
rhino	ay	r _ _ n _	This gray animal has a horn on its head. Use the shortened version of its name.
rhyme	ay	r _ _ _ _	Poems often do this. Some examples are seat and neat, break and take, moon and noon, or guess and mess.
ride	ay	r _ _ _	This is a verb that refers to the action you do with a rollercoaster, a car, or a horse.
rise	ay	r _ _ _	Yeast helps bread or dough do this. Also can be used to refer to waking up in the morning.
rye	ay	r _ _	This word may be used to refer to a bread. It may also be used to refer to whiskey made from this grain.
saliva	ay	s _ _ _ v _	This is another name for spit.
shy	ay	s _ _	This word may be used to refer to someone who is reserved. It has the same meaning as 'timid.'
side	ay	s _ _ _	A cube has six of these. This word can also refer to a less important dish served in addition to a main meal.
sign	ay	s _ _ n	Stop', 'one way', 'do not cross', and 'road work ahead' are all examples of this written mode of communication that often gives instructions.
sinus	ay	s _ n _ _	This word refers to cavities behind the nose and cheeks. You may feel pressure in these when you have a cold or infection.
size	ay	s _ _ _	To know if an article of clothing will fit, you should check this on the tag. It may be marked 'Small', 'Medium', or 'Large.'
sky	ay	s _ _	This is what we see when we look up in the air. It holds the clouds and is said to look blue.
slide	ay	s _ _ _ e	This word can refer to a piece of playground equipment, or a single page of a PowerPoint presentation.
sly	ay	s _ _	This word can mean tricky or clever. It is sometimes associated with a fox.
spider	ay	s _ _ _ _ _	This word refers to an 8-legged arachnid.
spine	ay	s p _ _ _	This part of the body runs through the back from the base of the neck to the tailbone.
spy	ay	s _ _	This word refers to a person who secretly collects information about the enemy. Other terms include mole, secret agent, or plant.
thigh	ay	t h _ _ _	This word refers to the upper part of the leg between the knee and the pelvic bone. It is usually considered part of the leg.
tide	ay	t _ _ _	This word is both a brand of laundry detergent and a term for a change in sea levels due to gravitational forces of the moon and sun.
tie	ay	t _ _	Men often wear one of these around their necks with suits.
time	ay	t _ _ _	This is what a clock or watch tells.
tribe	ay	t r _ _ _	This word refers to a group of people including Cherokee and Sioux. It is generally associated with Native Americans.
try	ay	t _ _	The saying goes, if at first you don't succeed, ______ again.
vine	ay	v _ _ _	Grapes grow on this.
vinyl	ay	v _ _ _ l	This is a material similar to plastic that may be used for wallpaper or flooring. It is also used to refer to musical records.
why	ay	w _ _	This word goes along with other question words such as Who, What, When, and Where.
wide	ay	w _ _ _	This word means the opposite of narrow. You may call a road with many lanes this W-word.
wine	ay	w _ _ _	This is an alcoholic beverage usually made from fermented grapes. Examples include Merlot or Cabernet.
wise	ay	w _ _ _	This word can mean 'smart' or 'clever,' and is often used to refer to elders or those with more experience. Wizards or owls may be associated with this word.

word	vowel	hint	clue
box	aa	b _ _	This is usually made of cardboard. You may mail or store things in one of these.
chalk	aa	c _ _ _ k	This is what is used to draw on a blackboard, for instance.
cough	aa	c _ u _ _	When you are sick, you may sneeze or do this.
fall	aa	f _ _ _	This is another word for the season 'Autumn.'
hog	aa	h _ _	This is another word for a 'pig.'
job	aa	j _ _	This is an informal name for your occupation or career.
plot	aa	p _ _ t	This word refers to the events that occur in a book or movie.
raw	aa	r _ _	This is the opposite of 'cooked.'
rock	aa	r _ _ _	Another word for 'stone.' Also a genre of music.
sock	aa	s _ _ _	This is an article of clothing you wear on your foot. It goes on before your shoes.
apple	ae	a _ _ _ e	This red fruit grows on trees.
bat	ae	b _ _	This is a nocturnal mammal with wings, or a tool used to play baseball.
cat	ae	c _ _	This household pet often meows.
crab	ae	c _ a _	Many people eat this sea creature, which is similar to a lobster.
fat	ae	f _ _	This word means the opposite of 'skinny.'
half	ae	h _ _ _	50 percent is equivalent to this amount.
lab	ae	L _ _	This is a sort of dog, or a place where an experiment could take place.
lap	ae	L _ _	This is the part of the body that you may have a child sit on.
last	ae	L _ _ _	This is the opposite of 'first.'
trash	ae	t _ _ _ _	This is another word for 'garbage.'
bed	eh	b _ _	This is the piece of furniture you sleep on. A blanket and pillow go on top.
deaf	eh	d _ _ _	This word refers to a person who is unable to hear.
depth	eh	d _ p _ _	You may use the word 'length' to refer to how long an object is. This word refers to how deep an object is.
desk	eh	d _ _ _	Schools and offices have this piece of furniture, which you may sit at to do work.
dress	eh	d _ _ _ _	A gown is another word for this article of clothing, which females usually wear.
head	eh	h _ _ _	This is the part of your body where you wear hats.
pet	eh	p _ _	This is a tame animal you may keep in your home, such as a cat, dog, or fish.
ten	eh	t _ _	This number comes before 11.
web	eh	w _ _	This is another name for the internet. Arachnids also spin these to catch their food in.
wrestle	eh	w _ _ _ _ _ _	This verb means to grapple with an opponent and then throw or force them to the ground. The WWE pays professionals to do this in the ring.
bait	ey	b _ i _	This word refers to a substance used to attract and catch fish. It may be a worm, another fish, or made of plastic.
bake	ey	b _ _ e	This refers to a way of cooking, which involves an oven. It is done when making cookies or cakes.
eight	ey	e _ _ _ _	This is an even number greater than seven. Also the product of two and four.
gray	ey	g _ _ _	This is the color most people's hair will turn as they age.
rain	ey	r _ _ _	This word refers to a form of precipitation. You may get wet, though umbrellas are useful for this.
ray	ey	r _ _	This word may refer to the beams of heat that seem to stream from the sun. It can be a male first name or an ocean creature, too.
table	ey	t _ _ _ _	This is a piece of common household furniture. It has four legs and you often eat at it.
tape	ey	t _ _ _	This word refers to an adhesive that comes in many forms such as packing, duct, electrical, and masking.
train	ey	t _ _ _ _	This is a mode of transportation with cabins and a caboose. It travels along a railway track.
wave	ey	w _ _ _	This can refer to the movement of one's hand to and fro to signal a greeting, or a long ripple of water that curls then breaks on the shore.
city	ih	c _ _ _	This word refers to a large settlement. For example, Detroit, Pittsburgh, or Philadelphia.
crib	ih	c _ _ b	Babies usually sleep in this. Similar to a cradle or bassinet.
fist	ih	f _ _ _	This is the shape your hand may take if you punch someone.
kid	ih	k _ _	This word may be used to refer to a young human or a baby goat.
kiss	ih	k _ _ _	This next word is an action you may do with your lips. Also a Hershey's candy.
lick	ih	L _ _ _	This action is done with your tongue. Dogs may also do this when they are happy to see you.
pig	ih	p _ _	This animal is known for rolling around in the mud and is where we get foods such as bacon and ham.
sick	ih	s _ _ _	This word means the same thing as 'ill.' You may take a day off of work or school if you have a cold and feel this way.
wrist	ih	w r _ _ _	This is the part of the body between the hand and the arm where you would wear a bracelet.
zip	ih	z _ _	This word may refer to a postal code or the action you perform to close up your jacket.

word	vowel	hint	clue
bee	iy	b _ _	This winged insect may not be as prevalent as we hope, but it is important for pollinating flowers and producing honey.
cheese	iy	c _ _ _ _ _	This is a dairy food. Examples are cheddar, Swiss, or mozzarella.
chief	iy	c h _ _ _	This next word may refer to a leader of a Native American group. Also, what the C in CEO stands for.
cream	iy	c _ _ _ _	People often drink their coffee with sugar and this dairy-based ingredient, similar to milk.
dream	iy	d _ _ a _	This word refers to a set of thoughts, images, and musings that may occur when you are asleep.
green	iy	g _ _ _ _	This is the color of most trees, grass, and other plants.
leap	iy	L _ _ _	This word means to jump, skip, and hop, too. It may also be associated with frogs.
pea	iy	p _ _	This small, round vegetable comes in a pod. It also plays a significant role in a Fairy Tale involving a princess who was kept awake because she could feel this vegetable, even though it was under many blankets.
treat	iy	t _ _ _ _	Before they get candy, children say the phrase ”trick or _____.”
wheat	iy	w _ _ a _	This is a grain. It’s part of the name of a well-known cereal, Shredded _____.
boat	ow	b _ _ _	This mode of transportation sails on water. Another word for ’ship.’
coat	ow	c _ _ _	This is an article of clothing that you may wear when it is cold out, such as in the winter. It is heavier than a jacket.
coma	ow	c _ _ _	This word refers to a state of deep unconsciousness due to illness or injury. People often go into or come out of these in soap operas.
comb	ow	c _ _ _	This next word is a grooming tool similar to a hair brush, but it has teeth rather than bristles.
nose	ow	n _ _ _	This is the part of the body most closely associated with our sense of smell. It is also useful for breathing.
pope	ow	p _ _ _	This next word refers to the position of the bishop of Rome and leader of the Roman Catholic church. The first name of the current one is Francis.
robe	ow	r _ _ _	This is clothing that you may wear after a bath or shower. It is long and loose and usually has a belt to keep it closed.
rope	ow	r _ _ _	This next word refers to a material that you can make into a knot or swing from. Lassos are made of it.
slow	ow	s _ _ _	This next word is used to mean the opposite of ’quick’ or ’fast.’
snow	ow	s _ _ _	This form of precipitation only occurs when it is cold enough. It comes down in flakes, usually during the winter.
drug	uh	d _ _ _	This may refer to an illicit substance like cocaine or marijuana, or a medication you can get at the pharmacy.
duck	uh	d _ _ _	This creature swims in lakes and ponds and is said to ’quack.’
luck	uh	L _ _ _	This is a term for good fortune. A four-leaf clover is said to bring this.
mud	uh	m _ _	This is what dirt turns into when it rains.
muffin	uh	m _ f _ _ _	This baked good is similar to a cupcake. It is baked in a tin and may be blueberry or banana nut.
mug	uh	m _ _	This is another word for a cup, such as one used for coffee or other hot drinks.
one	uh	o _ _	This number comes between zero and two.
rush	uh	r _ _ h	This word means to hurry. You don’t want to drive in ______ - hour traffic.
sub	uh	s _ _	This word can refer to a sandwich, an under water vessel, or an alternate teacher.
sun	uh	s _ _	This is what we call the star that the earth orbits.
boo	uw	b _ _	This word is something ghosts may utter when they attempt to scare you.
flu	uw	f _ _	You can get a shot each year to prevent this seasonal illness.
fruit	uw	f _ _ _ _	This is a category of food, similar to vegetables. Some examples are grapes, cantaloupes, watermelons, and berries.
lose	uw	L _ _ _	This is the opposite of ’win.’
nude	uw	n _ _ _	This word refers to wearing no clothing. It means the same as ’naked.’
rooster	uw	r _ _ _ _ _ _	This is a creature known for making loud calls early in the morning. Another word for a male chicken.
shoe	uw	s _ _ _	You wear this article of clothing on your foot when you leave your home.
snooze	uw	s _ o _ _ _	This word means ’sleep’. It is also a button on your alarm clock that you hit if you want to continue sleeping.
tuna	uw	t _ _ _	This is one of the most common fish to eat. When mixed with mayo, you can make it into a salad, similar to chicken salad.
zoo	uw	z _ _	This is a place where you pay to see animals such as zebras, elephants, and monkeys.

Appendix C: Lexical Decision Task Items

Fillers		Target Items	Non-Words
bang	pork	bribe	brap
bat	puke	buy	breel
bran	purse	chime	buke
burn	raft	chive	bune
cap	rash	climb	erlb
cheer	sand	crime	fram
church	smash	cry	fupe
class	stair	die	fute
clear	stale	dime	gamp
core	stamp	dive	gerp
corn	steer	drive	herg
cute	store	find	jamp
dare	strand	five	kune
deal	tail	fry	kupe
door	trail	glide	lork
fair	trans	grime	morsh
fang	trash	grind	muke
few	van	guy	nerlk
first	view	hind	peert
flap	weird	kind	porb
flirt		lime	pyoog
floor		live	quapp
fuel		mind	slail
fuse		mine	slan
gang		pint	snerp
hair		rhyme	snorp
ham		rind	spag
hang		rise	vort
hatch		shy	wearp
lamp		sign	werg
land		size	yair
last		slime	yan
laugh		sty	yand
learn		tie	yoosh
man		tribe	yorch
more		try	zaft
muse		twine	zamp
nurse		vine	zeer
orb		why	zerch
pair		wise	zerm

References

Aarts, H., Chartrand, T. L., Custers, R., Danner, U., Dik, G., and Jefferis, V. E. (2005). Social stereotypes and automatic goal pursuit. *Social Cognition*, 23:464–489.

Aarts, H., Custers, R., and Veltkamp, M. (2008). Goal priming and the affective-motivational route to nonconscious goal pursuit. *Social Cognition*, 26(5):555–577.

Aguilar, L., Downey, G., Krauss, R., Pardo, J., Lane, S., and Bolger, N. (2016). A dyadic perspective on speech accommodation and social connection: Both partners' rejection sensitivity matters. *Journal of Personality*, 84(2):165–177.

Arciuli, J. and Cupples, L. (2003). Effects of stress typicality during speeded grammatical classification. *Language and Speech*, 46(4):353–374.

Asch, S. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41:258–290.

Auer, P. and Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. In Auer, P., Hinskens, F., and Kerswill, P., editors, *Dialect change. The convergence and divergence of dialects in contemporary society*, pages 35–57. Cambridge University Press.

Babel, M. (2010). Dialect divergence and convergence in new zealand english. *Language in Society*, 39(4):437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40:177–189.

Babel, M., McGuire, G., Walters, S., and Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5(1).

Babel, M. E. (2009). *Phonetic and social selectivity in speech accommodation*. University of California, Berkeley.

Baese-Berk, M., Brent, T., Borrie, S., and McKee, M. (2015). Individual differences in perception of unfamiliar speech. In *Proceedings of the 18th International Congress of the Phonetic Sciences*. International Phonetics Association.

Bailey, G. (1997). When did southern american english begin? In Edward, S., editor, *Englishes around the world 1: Studies in honor of Manfred Gorlach*, pages 255–275. John Benjamines, Amsterdam.

Balcetis, E. and Dale, R. (2005). *An exploration of social modulation of syntactic priming*, pages 184–189. Lawrence Erlbaum Associates.

Bargh, J. (2006). What have we been priming all these years? on the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, 36(2):147–168.

Bargh, J., Chen, M., and Burrow, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71:230–244.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males, females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31:5–17.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Becker, K. (2016). Linking community coherence, individual coherence, and bricolage: The co-occurrence of (r), raised BOUGHT and raised BAD in New York City English. *Lingua*, 172–173:87–99.

Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2):145–204.

Bell, A. (2001). *Back in style: Reworking audience design*, page 139?169. Cambridge University Press, Cambridge.

Boersma, P. and Weenink, D. (2002). Praat, a system for doing phonetics by computer. *Glot international*, 5.

Bourhis, R. and Giles, H. (1977). *The Language of intergroup distinctiveness*. Academic Press, London, UK.

Brennan, S. E. and Metzing, C. A. (2004). Two steps forward, one step back: Partner-specific effects in a psychology of dialogue. *Behavioral and Brain Sciences*, 27.

Brown-Schmidt, S. and Tanenhaus, M. K. (2004). Priming and alignment: Mechanism or consequence? *Behavioral and Brain Sciences*, 27.

Brysbaert, M. and New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavioral Research Methods*, 41(4):977–90.

Byrne, D. (1971). *The attraction paradigm*. Academic Press, New York.

Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change*, 21:135–156.

Campbell-Kibler, K. (2011). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change*, 22:423–441.

Campbell-Kibler, K. (2016). *Toward a cognitively realistic model of meaningful sociolinguistic variation*.

Carey, B. (2011). Fraud case seen as a red flag for psychology research. *New York Times*.

Cebrian, J. (2006). Experience and the use of non-native duration in l2 vowel categorization. *Journal of Phonetics*, 34(3):372–387.

Cedergren, H. J. and Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 50(2):333–355.

Chartrand, T. and Bargh, J. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology*, 71(3):464–478.

Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press, Cambridge, MA.

Clopper, C. and Pisoni, D. (2007). Free classification of regional dialects of american english. *Journal of Phonetics*, 35(3):421–438.

Clopper, C. G. and Pisoni, D. B. (2005). Perception of dialect variation. In Pisoni, D. B. and Remez, R., editors, *The Handbook of Speech Perception*, chapter 13, pages 313–337. Blackwell.

Conrey, B., Potts, G., and Niedzielski, N. (2005). Effects of dialect on merger perception: Erp and behavioral correlates. *Brain and Language*, 95(3):435–449.

Crowne, D. P. and Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4):349–354.

Custers, R., Maas, M., Wildenbeest, M., and Aarts, H. (2008). Nonconscious goal pursuit and the surmounting of physical and social obstacles. *European Journal of Social Psychology*, 38:1013–1022.

Dijksterhuis, A. and Bargh, J. A. (2001). The perception–behavior expressway: Automatic effects of social perception on social behavior. *Advances in experimental social psychology*, 33:1–40.

Dimov, S., Katseff, S., and Johnson, K. (2012). Social and personality variables in compensation for altered auditory feedback. In Josep-Solé, M. and Recasens, D., editors, *The initiation of sound change: Perception, production, and social factors*. John Benjamins.

Dodsworth, R. (2014). Network embeddedness and the retreat from Southern vowels in Raleigh. *University of Pennsylvania Working Papers in Linguistics*, 20(2):41–50.

Dodsworth, R. and Kohn, M. (2012). Urban rejection of the vernacular: The SVS undone. *Language Variation and Change*, 24:221–245.

D'Onofrio, A. (2015). Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics*, 19(2):241–256.

Drager, K., Hay, J., and Walker, A. (2010). Pronounced rivalries: Attitudes and speech production. *Te Reo*, 53:27–53.

Earp, B. and Tramifow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Front. Psychol.*

Eckert, P. (2000). *Linguistic variation as social practice.* Blackwell, Oxford.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41:87–100.

Eisner, F. and McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Attention, Perception, & Psychophysics*, 67(2):224–238.

Eisner, F. and McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119(4):1950–1953.

Fasold, R. (1972). *Tense marking in Black English. A linguistic and social analysis.* Number 8. Urban Language Series.

Fischer, J. (1958). Social influence of a linguistic variant. *Word*, 14:47–56.

Fowler, C. (1986). Can event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14:3–28.

Fowler, C. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics*, 33:199–213.

Fowler, C. and Brancazio, L. (2000). Coarticulation resistance of american english consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech*, 43:1–41.

Fowler, C., Brown, J., Sabadini, L., and Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3):396–413.

Fridland, V. and Kendall, T. (2012). Exploring the relationship between production and perception in the mid front vowels of u.s. english. *Lingua*, 122(7):779–793.

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, 13(3):361–377.

Gambi, C. and Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, 4.

Gentilucci, M. and Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologia*, 45:608–615.

Giles, H. (1980). Accommodation theory: Some new directions. *York Papers in Linguistics*, 9(105–136).

Giles, H., Coupland, N., and Coupland, I. (1991). 1. accommodation theory: Communication, context, and. *Contexts of accommodation: Developments in applied sociolinguistics*, 1.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2):251–279.

Goldinger, S. D. and Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review*, 11(4):716–722.

Grondelaers, S. and van Hout, R. (2016). How (in)coherent can standard languages be? a perceptual perspective on co-variation. *Lingua*, 172-173:62–71.

Guy, G. R. (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics*, 52:63–71.

Guy, G. R. and Hinskens, F. (2016). Linguistic coherence: Systems, repertoires and speech communities. *Lingua*, 172–173:1–9.

Hall, J. S. (1942). The phonetics of great smoky mountain speech. *American Speech*, 17.

Hay, J. and Bresnan, J. (2006). Spoken syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review*, 23:321–349.

Hay, J. and Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4):865–892.

Hay, J., Nolan, A., and Drager, K. (2006a). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, 23(351–379).

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

Hedger, S. C. V., Heald, S. L., Koch, R., and Nusbaum, H. C. (2015). Auditory working memory predicts individual differences in absolute pitch learning. *Cognition*, 140:95–110.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1):B47–57.

Janse, E. and Newman, R. S. (2013). Identifying nonwords: effects of lexical neighborhoods, phonotactic probability, and listener characteristics. *Language and Speech*, 56(4):421–441.

John, O. P., D. E. M. and Kentle, R. L. (1991). *The Big Five Inventory–Versions 4a and 54*. Institute of Personality and Social Research, University of California, Berkeley.

Kim, D. and Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience*, 34(6):769–786.

Kim, M., Horton, W. S., and Bradlow, A. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2:125–156.

Kingston, J., Rich, S., Shen, A., and Sered, S. (2015). Is perception personal? In *Proceedings of the 18th International Congress of the Phonetic Sciences*. International Phonetics Association.

Koops, C., Gentry, E., and Pantos, A. (2008). The effect of perceived speaker age on the perception of pin and pen vowels in houston, texas. *UPenn Working Paper in Linguistics*, 14(2):Article 12.

Kraljic, T., Brennan, S. E., and Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107(1):54–81.

Kraljic, T. and Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology*, 51(2):141–178.

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

Krauss, R. M. and Pardo, J. S. (2004). Is alignment always the result of automatic priming? *Behavioral and Brain Sciences*, 27.

Kuznetsova, A., Brockhoff, P., and Christensen, R. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software, Articles*, 82(13):1–26.

Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3):273–309.

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 45(4):715–762.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (1994). *Principles of Linguistic Change, Vol. 1. Internal Factors.* Blackwell, Oxford.

Labov, W. (2006 [1966]). *The social stratification of English in New York City*. Cambridge University Press.

Labov, W. (2010). *Principles of Linguistic Change, Vol. 3. Cognitive and Cultural Factors.* Wiley-Blackwell.

Labov, W., Ash, S., and Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Mouton de Gruyter, Berlin.

Labov, W., Yaeger, M., and Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress*. Number v. 2 in A Quantitative Study of Sound Change in Progress. U.S. Regional Survey.

Lambert, W., Hodgson, R., Gardner, R., and Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, 60(1):44–51.

Lehet, M. and Holt, L. L. (2017). Dimension-based statistical learning effects both speech perception and production. *Cognitive Science*, 41(S4):885–912.

Lev-Ari, S. and Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics*, 47:36–46.

Lewandowski, N. and Jilka, M. (2019). Phonetic convergence, language talent, personality, and attention. *Frontiers in Communication*, 4(18):1–19.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36.

Lipski, S., Escudero, P., and Benders, T. (2012). Language experience modulates weighting of acoustic cues for vowel perception: An event-related potential study. *Psychophysiology*, 49(5).

Lobanov, B. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, 68:1636–1642.

Love, J. and Walker, A. (2013). Contextual activation of australia can affect new zealanders' vowel productions. *Journal of Phonetics*, 48:76–95.

Maye, J., Aslin, R. N., and Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3):543–562.

McGowan, K. (2015). Social expectation improves speech perception in noise. *Language and Speech*, 58(4):502–521.

McQueen, J. M., Cutler, A., and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6):1113–1126.

Mendoza-Denton, N. (2008). *Homegirls : Language and Cultural Practice among Latina Youth Gangs*. Blackwell, Malden, MA.

Miller, R. M., Sanchez, K., and Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception and Psychophysics*, 72(6):1614–25.

Mitterer, H. and Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: evidence from the shadowing task. *Cognition*, 109(1):168–73.

Namy, L., Nygaard, L., and Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21(4):422–432.

Natale, M. (1975a). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5):790–804.

Natale, M. (1975b). Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills*, 40:827–830.

Natale, M. (1976). Need for social approval as related to speech interruption in dyadic communication. *Perceptual and Motor Skills*, 42:455–458.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1):62–85.

Nielsen, K. (2011). Specificity and abstractness of vot imitation. *Journal of Phonetics*, 39(2):132–142.

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

Pardo, J., Gibbons, R., Suppes, A., and Krauss, R. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, 40(1):190–197.

Pardo, J., Jordan, K., Mallari, R., Scanlon, C., and Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, 69(3):183–195.

Pardo, J., Urmanche, A., Wilman, S., and Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception and Psychophysics*, 79(2):637–659.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119(4):2382–2393.

Pardo, J. S., Jay, I. C., and Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, 72(8):2254–2264.

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., and Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics*, 69:1–11.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., and Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America*, 130(1):461–472.

Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226.

Pickering, M. J. and Garrod, S. (2013). Uncovering the roles of gender stereotypes in speech perception. *Behavioral and Brain Sciences*, 36:329–392.

Pierrehumbert, J. (2001). *Exemplar dynamics: Word frequency, lenition and contrast*, pages 137–158. John Benjamins, Amsterdam and Philadelphia.

Podlipský, V. and Šimáčková (2015). Phonetic imitation is not conditioned by preservation of phonological contrast but by perceptual salience. *Proceedings of ICPhS 2015*.

Purnell, T., Idsardi, W., and Baugh, J. (1999). Perceptual and phonetic experiments on american english dialect identification. *Journal of Language and Social Psychology*, 18(1):10–30.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reed, P. (2014). Inter- and intra-generational monophthongization and southern appalachian identity. *Southern Journal of Linguistics*, 38:159– 193.

Reed, P. (2016). Sounding appalachian: /ai/ monophthongization, rising pitch accents, and rootedness. *Doctoral Dissertation*.

Reinisch, E. and Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2):539–555.

Rickford, J. R. and McNair-Knox, F. (1994). Addressee- and topic-influenced style shift: A quantitative sociolinguistic study. In Biber, D. and Finegan, E., editors, *Sociolinguistic Perspectives on Register*, pages 235–276. Oxford University Press.

Sanchez, K., Hay, J., and Nilson, E. (2015). Football versus football: Effect of topic on /r/ realization in american and english sports fans. *Language and Speech*, 56(4):443–460.

Sancier, M. L. and Fowler, C. A. (1997). Gestural drift in a bilingual speaker of brazilian portuguese and english. *Journal of Phonetics*, 25(4):421–436.

Schertz, J. and Clare, E. J. (2019). Phonetic cue weighting in perception and production. *Cognitive Science*, 11(2):1–24.

Staum Casasanto, L. (2010). What do listeners know about sociolinguistic variation? *University of Pennsylvania Working Papers in Linguistics*, 15(2):Article 6.

Stewart, M. E. and Ota, M. (2008). Lexical effects on speech perception in individuals with "autistic" traits. *Cognition*, 109:157–162.

Strand, E. (1999). Uncovering the roles of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, 18(1):86–99.

Sumner, M. and Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60(4):487–501.

Tamminga, M. (2019). Interspeaker covariation in philadelphia vowel changes. *Language Variation & Change*, 31(2):119–133.

Tamminga, M., Wade, L., and Lai, W. (2019). The search for predictors of individual differences in vot imitation. Paper presented at NWAV 48.

Theodore, R. M. and Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America*, 128(4):2090–9.

Thomas, E. R. (2011). Sociolinguistic variables and cognition. *Advanced review*.

Torbert, B. (2010). The salience of two southern vowel variants: Fronted /o/ and weak-glided /ai/. *Southern Journal of Linguistics*, 34(2):1–36.

Turnbull, R. (2015). Patterns of individual differences in reduction: Implications for listener-oriented theories. In *Proceedings of the 18th International Congress of the Phonetic Sciences*. International Phonetics Association.

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4):580–591.

Wade, L. (2017a). Do speakers converge toward variants they haven't heard? Paper presented at NWAV 46.

Wade, L. (2017b). The role of duration in the perception of vowel merger. *Laboratory Phonology*, 8(1):30.

Wade, L. and Roberts, G. (2020). Linguistic convergence to observed versus expected behavior in an alien language map task. *Cognitive Science*, 44(4).

Walker, A. and Campbell-Kibler, K. (2015). Repeat what after whom? exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, 6.

Walker, A., Hay, J., Drager, K., and Sanchez, K. (2018). Divergence in speech perception. *Linguistics*, 56(1).

Warren, P., Rae, M., and Hay, J. (2003). *Word recognition and sound merger: The case of the front-centering dipthongs in NZ English*, pages 2989–92. Rundle Mall: Causal Productions, Barcelona.

Weatherholtz, K. (2015). *Perceptual learning of systemic cross-category vowel variation*. PhD thesis, The Ohio State University.

Weinreich, U., Labov, W., and Herzog, M. I. (1968). Empirical foundations for a theory of language change. In Lehmann, W. and Malkiel, Y., editors, *Directions for historical linguistics: A symposium*, pages 97–195. University of Texas Press, Austin and London.

Wilson, C., Chodroff, E., and Nielsen, K. (2016). Generalization in vot imitation: Feature adaptation or acoustic-phonetic covariation? *Acoustical Society of America*.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. B., and Wapner, S. (1954). *Personality through perception: an experimental and clinical study*. Harper.

Wolfram, W. and Christian, D. (1976). *Appalachian Speech*. Center for Applied Linguistics, Arlington, VA.

Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS One*, 5(8):e11950.

Yu, A. C. L., Abrego-Collier, C., and Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE*, 8(9):e74746.

Yuan, J. and Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics 2008*, pages 5687–5690.

Zehr, J. and Schwarz, F. (2018). Penncontroller for internet based experiments (ibex).

Zellou, G., Dahan, D., and Embick, D. (2017). Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition And Neuroscience*, page 1? 16.

Zellou, G., Scarborough, R. A., and Nielsen, K. (2013). Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America*, 140(5):3560–75.