EXPERIMENTAL EVIDENCE FOR EXPECTATION-DRIVEN LINGUISTIC CONVERGENCE

LACEY WADE

University of Pennsylvania

This article examines the role of sociolinguistic expectations in linguistic convergence, using glide-weakened /aɪ/—a salient feature of Southern US English—as a test case. I present the results of two experiments utilizing a novel experimental paradigm for eliciting convergence—the WORD-NAMING GAME task—in which participants read aloud (baseline) or hear (exposure) clues describing particular words and then give their guesses out loud. Participants converged toward a Southern-shifted model talker by producing more glide-weakened tokens of /aɪ/, without ever hearing the model talker produce this vowel. Participants in the control (Midland talker) condition exhibited no such response. Convergence was facilitated by both living in the South and producing less-weakened baseline /aɪ/ glides, but attitudinal and domain-general individual-differences measures did not reliably predict convergence behaviors. Results are discussed in terms of implications for the cognitive mechanisms underlying convergence behaviors and the mental representations of sociolinguistic knowledge.*

Keywords: sociolinguistic cognition, convergence, individual differences, experimental sociolinguistics, Southern US speech, glide-weakened /aɪ/

1. INTRODUCTION. Work in sociolinguistics has long shown that social and linguistic information are linked. Not only does linguistic variation pattern along social dimensions and index social meaning, but language users also possess implicit sociolinguistic knowledge that influences the way language and people are perceived. As such, recent work in sociolinguistic cognition has shed light on the way sociolinguistic knowledge is mentally represented and utilized—usually in speech perception (Thomas 2011, Loudermilk 2015, Campbell-Kibler 2016). This article adds to the growing body of knowledge on sociolinguistic cognition by examining language users' sociolinguistic expectations, with the goal of shedding light on the way sociolinguistic knowledge is mentally represented and utilized in speech production specifically. LINGUISTIC CONVERGENCE—in which a language user adjusts their speech to become more similar to an interlocutor—offers an excellent opportunity to work toward this goal. Convergence[1] 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). Convergence is an ideal test case for probing questions about the relationship between social and linguistic knowledge because it has been reported to occur in response to both linguistic and social cues, and it may have broader implications for many important questions in linguis-

* This work was funded by NSF grant BCS-1917900. I'd like to thank Andries Coetzee, Lauren Hall-Lew, Abby Walker, and an anonymous referee for thoughtful feedback that ultimately improved this paper. Thanks also to project RAs Sadie Butcher and Leila Pearlman for their tireless work editing TextGrids, to Walt Wolfram for helping me recruit participants at NCSU and letting me borrow his office for the week, and to Meredith Tamminga, Gareth Roberts, David Embick, and members of the Language Variation and Cognition lab at Penn for feedback at various stages of this project. All remaining errors are of course my own.

[1] This process has many different names within linguistics, social psychology, and cognitive science, including *alignment*, *spontaneous imitation*, and *entrainment*. While the term *convergence* is also often used in the context of language contact, 'language convergence' is sometimes thought of as a large-scale, longer-term instantiation of the type of interpersonal convergence observed in short-term interactions. Indeed, this is essentially what a 'change-by-accommodation' model (e.g. Trudgill 1986) proposes.

tics, including the relationship between perception and production and the mechanisms behind language change.

Here I present the results of two experiments designed to test whether individuals converge toward a variant they may reasonably expect, but do not directly observe, from a model talker. Specifically, I ask whether speakers imitate a Southern-shifted model talker by producing more glide-weakened tokens of the /aɪ/ vowel—a salient feature of Southern US English—without ever actually hearing this vowel from the model talker. With this goal in mind, I first propose a terminological and theoretical distinction differentiating two types of convergence that vary primarily in the relationship between the TRIGGER—any cue that initiates convergence—and the TARGET—the linguistic form that is intended during convergence. Most work on convergence focuses on a linguistic target that is triggered by observing the same linguistic form in real time.[2] An example would be producing lengthened voice onset time (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 immediate input.

I contrast this with what I refer to here as EXPECTATION-DRIVEN CONVERGENCE, which entails converging toward a target that was encountered at some point in the past—not necessarily from the same talker or in the same context—and is later recalled but not locally observed. The trigger of convergence in these cases can be linguistic, but need not be,[3] and the target is derived from expectations stemming from preexisting 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 their Southern interlocutor 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,[4] in part because existing experimental paradigms such as shadowing tasks—where participants repeat words aloud after a model talker—are not well suited to elicit expectation-driven convergence. Two major goals of this article are therefore (i) to determine whether expectation-driven convergence occurs in a controlled experimental setting and, in order to do this, (ii) to develop and test a suitable experimental paradigm for eliciting expectation-driven convergence.

1.1. THE ROLE OF EXPECTATION AND GENERALIZATION IN CONVERGENCE. The crucial difference between input-driven 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 maximally similar. For instance, an individual may hear their interlocutor produce *pancake* with lengthened VOT on /p/ and then produce the word *pancake* with lengthened VOT on /p/ themself.

[2] *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, and 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.

[3] For instance, a cue to Southernness may be a speaker's attire (such as wearing a jersey of a Southern sports team) or the interaction actually taking place in the South.

[4] There has been experimental evidence for expectation-driven shifts in production, often framed in terms of topic-based style shifting, which may very well be driven by the same mechanisms of expectation-driven convergence. However, I interpret expectation-driven convergence to refer specifically to shifting one's speech to become more similar to the expected linguistic behavior OF AN INTERLOCUTOR.

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 a lengthened-VOT /p/, or perhaps generalize even further and produce *kitten* with a lengthened-VOT /k/. Of course, some degree of generalization from the target to the trigger is always necessary, even in the most maximally input-driven instances of convergence, because the context in which a word is perceived is nonidentical to the context in which it is later spoken, and some degree of generalization from interlocutor to self is required during convergence. 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 phoneme with shared features? The trigger may even be so far removed from the linguistic target as not to be linguistic at all. In theory, varying degrees of generalization are possible, and the question of the extent to which convergence generalizes is not new in convergence research. If we assume that the trigger and target must share a representation at some level, we may also frame this question as 'What types of representations are utilized in convergence?'.

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. One proposal is that the target of convergence is quite narrow and unable to generalize. For instance, Goldinger (1998) and 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 heard, but the 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 those of 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 generalize across phonemes. Nielsen (2011) found that when participants were exposed to artificially lengthened VOT for /p/-initial words, not only did they 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. Wilson et al. (2016) suggest that the speakers may have produced lengthened VOT in /k/-initial words after hearing lengthened VOT in /p/-initial words because they were accessing knowledge that VOT of /k/ and /p/ generally covary within speakers. Indeed, Theodore and Miller (2010) show that listeners know that longer VOT for /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/. Such work brings up an intriguing question about what we are actually mentally representing about INDIVIDUAL SPEAKERS.

Zellou et al. (2017) also propose 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 hypernasalized speaker in the first block of shadowing increased their degree of coarticulatory nasalization. However, if hypernasalization 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 across all tokens they had heard from the speaker and converging toward that

average. Zellou et al. 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. Such findings have provided evidence that convergence may target broader representations than previously thought.

While automatic, mechanistic accounts have tended to focus on input-driven convergence, social-psychological accounts that appeal to motivations such as identity and managing social distance (e.g. COMMUNICATION ACCOMMODATION THEORY; Giles et al. 1991) have allowed for a wider range of convergence behaviors that might fall under the label of 'expectation-driven convergence'. For instance, Auer and Hinskens (2005: 343) advocate for an IDENTITY-PROJECTION MODEL of convergence, which 'does not mean imitating the actual speech of one's co-participant, but rather conforming to some stereotyped image of how a person in the social role of the co-participant ought to, or can be expected to, behave'. A somewhat anecdotal example comes from Bell (2001), of a male Māori interviewee using the most *eh*-tags when conversing with a male Māori interviewer who used hardly any, and an Anglo interviewer using the most *eh*-tags when conversing with a Māori interviewee who used far fewer, both attributed to the fact that *eh*-tags are a stereotypical feature associated with male Māori speech. The idea that convergence is triggered by assumptions about the speaker rather than by the speech itself can be traced further back to Thakerar et al. (1982), who define accommodation that 'responds to what the speaker mistakenly assumes will be the addressee's speech on the basis of the addressee's nonspeech attributes' as 'subjective accommodation' (Bell 1984:168).

Bell's (1984) AUDIENCE DESIGN model of style shifting also predicts expectation-driven speech adjustments. The audience design model assumes that when people speak, they take into account primarily those who will be listening and design their speech accordingly. Speech may be influenced not only by one's primary addressee, but also by others who may be involved in the communicative situation (auditors, overhearers, eavesdroppers), as well as by other nonpersonal variables such as topic and setting. While there are a number of ways that an individual may shift their speech when taking those around them into account, perhaps the most obvious is converging TOWARD the linguistic behaviors observed (from interlocutors) or expected (from auditors, overhearers, or eavesdroppers), the latter of which are likely instances of expectation-driven convergence.[5] Other findings in the realm of sociolinguistic style shifting might also be considered instances of expectation-driven convergence. For instance, in Fasold's (1972) work on African American English in Washington, DC, he observed that speakers produced more African American English (AAE) features such as *d*-deletion when conversing with an African American interviewer than with a White interviewer, though most of the interviewers were middle-class women who spoke Mainstream American English rather than AAE. Similarly, Rickford and McNair-Knox (1994) found that an African American teenager 'Foxy Boston' used more AAE features when conversing with African American interviewers, and also when talking about friends, boys, and dating compared to when discussing school.

Experimental evidence for topic-based style shifting has also shown that speakers shift their speech toward features they do not directly observe. For instance, Drager et al. (2010) found that New Zealanders shifted their pronunciation in a word-list reading

[5] Thanks to Lauren-Hall Lew for pointing out that shifts in speech directed toward children, nonnative speakers of English, pets, and AI (like Siri or Alexa) are often expectation-driven as well.

task to incorporate more Australian-like vowels after reading facts about Australia. Similarly, Love and Walker (2013) found that British and American soccer fans became more /r/-ful when talking about American football than when talking about soccer, which they suggest is because /r/-fulness is associated with American English. Walker (2019) later replicated this finding, showing that the topics of US government and US football triggered both more /r/-fulness and more /t/-flapping than the topics of UK government or soccer. Importantly, these effects persisted even for non-topic-specific items, suggesting that the effect was not driven by topic-specific lexemes. Relatedly, Sanchez et al. (2015) found that New Zealand speakers in a corpus produced more Australian-like variants of the KIT and TRAP vowels when talking about Australian topics, and experimental participants showed the same effect for the DRESS vowel in the context of producing Australia-related lexical items. While the relationship between style shifting and convergence remains largely unclear, there is much overlap in the types of behavior that could be considered style shifting or convergence. Even though topic-based shifting does not stem from interlocutor-directed expectations, evidence that speakers shift their speech to include particular variants that they do not directly observe in that moment seems to support the kinds of cognitive mechanisms that would also make expectation-driven convergence possible.

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 that 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 the 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 in natural language.

1.2. THE PRESENT STUDY. The present study focuses on US Southern speech as a test case in probing expectation-driven convergence, with the broader goal of shedding light on the cognitive mechanisms underlying linguistic convergence and the mental relationship between social and linguistic knowledge. I investigate the role of expectations regarding 'Southernness' and glide-weakened /aɪ/—a salient feature of US Southern English—in convergence behaviors. I present the results of two experiments testing whether participants shift their speech to produce more glide-weakened tokens of /aɪ/ when conversing with a Southern-shifted model talker who crucially never produces the /aɪ/ vowel. Both experiments utilize a novel WORD-NAMING GAME paradigm designed 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 to reduce potentially inhibitive effects of speech monitoring.

The Southern US dialect and /aɪ/ glide-weakening specifically were chosen for this study because of their 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. The South is perhaps the most distinct dialect region in the US, and glide weakening of /aɪ/ is a particularly salient and stereo-

typical feature of Southern English[6] (e.g. Hall 1942, Wolfram & Christian 1976, Reed 2014). As Reed (2016:17) suggests, 'American English speakers are aware of monophthongization and its status as a regional and subregional linguistic caricature … Virtually every popular depiction of Southern and Appalachian speech displays monophthongal /aɪ/ as a noteworthy feature'. Reed goes on to note 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 was a salient perceptual cue indexing 'Southernness', though another Southern dialect feature, the fronting of /o/, was not. Though there has not been much work 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 (2010:55) notes, '[t]he most generally recognized feature of Southern speech is the monophthongization of /aɪ/'.

Weakening of the /aɪ/ glide is often described as the triggering event of the 'Southern Shift', a vocalic chain shift affecting practically the entire vowel space. 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 & Kohn 2012, Dodsworth 2014). The glide-weakened, Southern-shifted realization of /aɪ/ is closer to [aː], with a glide lower and more central in the vowel space such that there is little to no movement from nucleus to glide (Labov 1994). The phonological constraints and social stratification of this shift vary throughout the South: in more inland regions of the South, 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 occurs only in coda position or preceding voiced segments.

Experiment 1 finds evidence for expectation-driven convergence toward Southern speech in the lab, and experiment 2 replicates this finding online with a different subject pool and new model talkers. In addition to establishing expectation-driven convergence as a real and replicable phenomenon, the experiments presented here also investigate some of the factors that may facilitate this behavior, including dialect experience, cognitive/personality traits, and social attitudes. Understanding how both situational variables, such as attitudes toward a model talker, and stable individual traits, such as cognitive style, impact the likelihood and degree to which an individual will converge may point toward the cognitive processes utilized in convergence behaviors generally.

There are several specific research questions to be addressed:

(i) Do participants converge toward a variant they may reasonably expect—but do not observe—from a model talker?
(ii) If so, how long does convergence last post-exposure?
(iii) How does sociolinguistic experience and/or dialect background mediate convergence?
(iv) Do individuals conform to the aggregate group pattern? If not, which individual factors (both stable and situational) predict convergence rates?

2. METHODS. Linguistic convergence is a well-established phenomenon that has been observed both in naturalistic, dyadic settings and 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).

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

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 are not applicable to 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 in order 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 must 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 target tokens is a nontrivial task. More importantly, it would be extremely difficult to ensure that 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.

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 in order 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. This novel experimental paradigm was first used on a smaller scale in experiment 1 in order to test that it sufficiently elicits convergence behavior, and minor modifications were then made to it between experiments 1 and 2, mainly to elicit more data. The methods used in both experiment 1 and experiment 2 are described below, with the few differences between the experiments highlighted.

2.1. Procedure.

Set-up. Participants completed experiment 1 in person on a laptop computer and were recorded with a Yeti microphone. The experiment was programmed and administered in PsychoPy. Experiment 2 was administered online in order to recruit participants more quickly and from a wider variety of regions within the US, and it was programmed in Ibex (Alex Drummond, https://adrummond.net/ibexfarm) using the PennController (Zehr & Schwarz 2018). This version was completed in an internet browser, and participants' voices were recorded through their computer microphones.[7] In both experiments, after giving informed consent, participants tested their microphone and headphones. They then completed a brief demographic questionnaire, which collected information on their age, gender identity, and residential history, with race/ethnicity and

[7] Scripts and sample stimuli used in the design of this experiment are available at https://osf.io/ja6vy/?view_only=56f3fcb658e64d9b832aae8fbbf64fe4.

education information additionally collected for experiment 2. Residential history was particularly important for labeling participants as 'Southern' or 'non-Southern'.

WORD-NAMING GAME. Participants then completed the three phases of the word-naming game. In this task, 'clues' were presented to participants in written or audio format, depending on the experimental phase, and participants named aloud into the microphone the word described by the clues, using the carrier phrase: 'The word is __ '. For instance, a participant may receive 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"'. In experiment 1, participants were recorded for the duration of the experiment. In experiment 2, participants pressed a 'Record' button on the screen, the screen indicated that they were recording, and then they gave their response. When they finished recording their response, participants pressed the 'Next' button to continue to the next clue, and the screen indicated that they were no longer recording. In an attempt to minimize data loss due to incorrect responses, hints were also provided. In experiment 1, participants were allowed to ask for a hint by pressing the 'h' key on their keyboard at any point throughout the course of the experiment. In experiment 2, participants were automatically provided with a hint at the bottom of the screen for all clues. Hints contained spaces indicating the number of letters in a word, with one to three letters filled in (e.g. *t* _ _ *r*). An example of what the screen looked like during a trial in experiment 2 is shown in Figure 1.

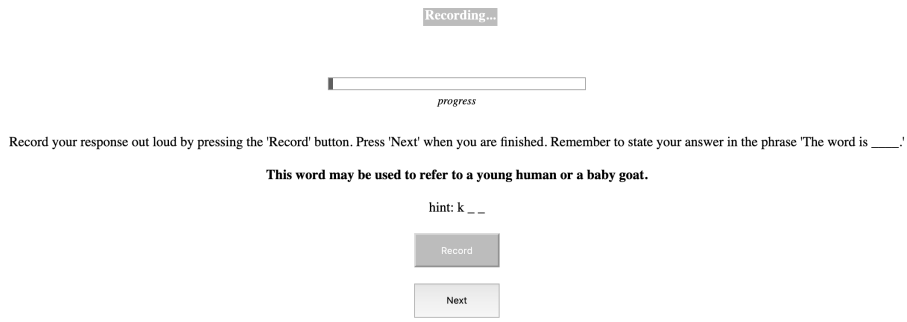

FIGURE 1. Screenshot of a trial in the baseline/post-exposure phase of experiment 2, where clues are presented on screen.

The word-naming game 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 itself was the same. Clues in this phase were read 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 a 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 resulting from various factors. For instance, since this experiment targets /aɪ/ glide-weakening, which is a reduction process, a result of general fatigue as the experiment progresses (i.e. more re-

duced 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. The 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.

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 but over the headphones this time so that participants could familiarize themselves with the model talker. The auditory instructions were somewhat modified from the earlier orthographically presented instructions, partially in order 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 auditory directions.

Elicited tokens consisted of target /aɪ/ words in word-final or pre-voiced-consonant contexts, as this is the environment in which /aɪ/ glide-weakening reliably occurs in the US South. In experiment 1, sixty target words and sixty non-/aɪ/ filler words were elicited. In order to collect more data per participant per phase, experiment 2 elicited ninety target tokens (including the sixty elicited in experiment 1). Ninety non-/aɪ/ filler words were also elicited. Tokens were elicited in three sets, each consisting of an equal number of /aɪ/ words and fillers, with one set being elicited per 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 & New 2009) Log10CD measure. For each set of target items, lexical frequency had the same mean and standard deviation.

Clues were one to two sentences long and gave participants some evidence for the dialect of the talker. A complete list of the clues used to elicit the items is given in the online supplementary materials.[8] Crucially, as noted above, the /aɪ/ vowel was never included in any of the clues. When clues were presented auditorily, they were read by a young, white talker who was from either the Midland or the South. Both talkers in experiment 1 were male, and both talkers in experiment 2 were female. While this decision was made primarily based on model-talker availability, the change in talkers also serves to test replication of the main finding with model talkers of a different gender

[8] The supplementary materials can be accessed at http://muse.jhu.edu/resolve/144.

and from a different region of the South. The Southern talker in experiment 1 was from western North Carolina, and the one in experiment 2 was from Hurley, Mississippi, and both produced typical Southern dialect features such as raised front lax vowel nuclei, fronting of back vowels, and the PIN-PEN merger. The Midland talkers were both from northeast Ohio and were described by participants as 'normal' or 'neutral' sounding. Model-talker vowel plots are provided in Appendix A.

ADDITION OF POST-TASK SURVEY. Experiment 2 added a follow-up Qualtrics survey after the word-naming game, intended to assess the role of individual differences in convergence behaviors. The survey consisted of the following.

- General demographic information and questions regarding what participants believed the experiment to be about and the perceived regional identity of the talker
- Attitudinal questions about the model talker: participants were asked to rate the model talker they heard on twelve personal attributes, using a slider bar with values ranging from 0–100 (e.g. 'The talker sounded kind'). Five of these are combined for a 'Perceived talker prestige' score, and five others are combined for a 'Perceived talker likability' score.
- Attitudinal questions about the South: participants were given nine statements about the South and were asked to determine the extent to which they agreed with each, using a slider bar ranging from 0–100 (e.g. 'I am familiar with the way Southerners speak'). Seven of these gauge familiarity with the South and are summed for a combined 'Familiarity with the South' score. The statements that participants rated for these sections are provided in the online supplementary materials.
- The BIG FIVE INVENTORY (John et al. 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. Participants rated the extent to which they agreed with forty-four statements on a five-point Likert scale.
- The AUTISM-SPECTRUM QUOTIENT (AQ; 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). Subcategories measured by the AQ include attention switching, social skills, imagination, communication, and attention to detail. Participants rate the extent to which they agree with fifty statements (ten for each subscale) on a four-point Likert scale, and higher scores in each of these categories indicate more 'autistic-like' traits.
- The MARLOWE-CROWNE SOCIAL DESIRABILITY SCALE (Crowne & Marlowe 1960): a self-reported measure indicating the extent to which an individual misrepresents themself to earn the approval of others. The survey consists of statements that are unlikely to be entirely true, but reflect socially desirable traits (e.g. 'I have never been irked when people express ideas different from my own'), and it is often used to determine how truthful an individual will be on other self-assessment surveys. The survey consists of thirty-three true/false responses, which are scored on a scale of 0–33.

These various cognitive-style/personality measures were added in experiment 2 for two main reasons. The first is to test whether the individual-differences measures previously shown to predict degree of input-driven convergence also predict degree of EXPECTATION-DRIVEN CONVERGENCE in the present study. The second is to determine which measures predict expectation-driven convergence, which (when compared to predictors

for input-driven convergence and other linguistic behaviors) may shed some light on the mechanisms underlying the behavior.

2.2. Analysis. Recordings were forced-aligned using the Penn Phonetics Lab Forced Aligner (P2FA) (Yuan & 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, including hand-adjusting all TextGrids.[9] Formant-tracking settings in Praat were adjusted for each speaker, with defaults of five formants tracked and a max formant height of 5500 Hz for females and 5000 Hz for males. A script was run in Praat (Boersma & Weenink 2002) to automatically measure the first three formants at seven 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), and all data visualization and statistical modeling were done in R (R Core Team 2020). The primary measure used in the following analysis is the phonetic realization of the offglide (at 80%) along the front diagonal of the vowel space, measured as normalized F2 – 2 * F1, following Dinkin and Dodsworth (2017). This measure captures the peripherality of the glide, such that a higher value indicates a fronter and higher realization. The diagonal measure was chosen (as opposed to e.g. Euclidean distance) in order to isolate the effects of the experimental manipulation without introducing additional variability from general undershoot (i.e. due to fatigue or acclimation to the task) as the experiment progressed, which was observed in nucleus measures in the control condition. The diagonal measure was deemed more appropriate in that it remained most stable across the course of the control condition, where we would not expect shifts in /aɪ/ realization.

Inaccurate responses and tokens with sound quality that was so poor that they could not be acoustically analyzed were omitted (2.6% of tokens in experiment 1 and 5.1% in experiment 2). Outliers were excluded from the following analyses on a by-item and by-participant basis, with observations greater than three standard deviations from the mean being omitted, accounting for less than 1% of tokens in either experiment. For experiment 1, this leaves a total of 4,607 /aɪ/ tokens (or roughly fifty-eight per person, and nineteen per person per experiment phase) in the analysis. For experiment 2, a total of 9,491 /aɪ/ tokens (or roughly eighty-four per person, and twenty-eight per person per experiment phase) were analyzed.

In both experiments, a linear mixed-effects regression model was fit to the data using the lmerTest (Kuznetsova et al. 2017) package in R (R Core Team 2020), which provides *p*-values for lmer model fits using Satterthwaite's degrees of freedom method. Front diagonal (F2 – 2 * F1) at the glide (80%) was the dependent variable. All fixed effects available and of interest to the research questions of the study were included, and are listed below.

- **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.

[9] Two RAs and the author hand-adjusted TextGrids blindly and in a random order, such that TextGrid editors were not aware of the condition, phase, and order that a sound file came from while editing.

- **Baseline:** Continuous predictor referring to the average baseline diagonal measure of each participant, *z*-scored.
- **Dialect:** Categorical predictor referring to a participant's dialect background, with levels SOUTHERN (reference) and NON-SOUTHERN, sum coded.
- **Duration:** Continuous predictor referring to token duration, *z*-scored.
- **Frequency:** Continuous predictor referring to lexical frequency, *z*-scored.
- **Phase * Voice * Dialect:** Three-way interaction between Phase, Voice, and Dialect. Note that Dialect is sum coded so that the main effects of Phase (and Voice) refer to both dialect groups.
- **Phase * Voice * Baseline:** Three-way interaction between Phase, Voice, and Baseline. Note that 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 to only the Southern voice condition, and the main effect of Voice is restricted to only the baseline phase.

For experiment 1, random intercepts are included for subject and item. No random slopes are included due to lack of convergence for any random-slopes models, which partially motivated the methodological changes in experiment 2 designed to increase token count and therefore statistical power. The final model for experiment 1 is shown in Table 1 below. For experiment 2, random intercepts for subject and item are again included. All fixed predictors and interactions were tested as random slopes where appropriate, and those that significantly improved the model based on likelihood-ratio test significance are included. This includes by-participant random slopes for Phase, Duration, and Frequency and by-item random intercepts for Duration and Baseline. The summary of fixed effects for the final model is shown in Table 2 below. Data and scripts are available at https://osf.io/2tkr8/?view_only=4f4f86f94ab54a60b9462abc38560499.

2.3. PARTICIPANTS.

EXPERIMENT 1. Eighty-four participants from either the University of Pennsylvania or North Carolina State University were recruited for experiment 1. Participants were labeled as 'Southern' or 'non-Southern', depending on their residential history. Participants who spent the majority of their school-aged years (ages five to eighteen) within the South were labeled as 'Southern'. Forty-two participants, thirty-nine of whom 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 participants fell into this category, all of whom were recruited from the University of Pennsylvania. All participants were 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 and technical issues with the participant's recording. Data is therefore analyzed for forty Southern participants and thirty-nine non-Southern participants. Of these seventy-nine participants, fifty-nine self-identified as female, nineteen as male, and one as nonbinary. Mean age is 20.75 (range: 18–42). All participants were given either course credit or a $10 Amazon gift card in exchange for their participation.

EXPERIMENT 2. For experiment 2, 120 participants were recruited to participate. Of these, 105 were recruited through Prolific and completed the experiment online for $10. Fifteen were recruited through the University of Pennsylvania subject pool and received course credit for completing the study in person at the Language Variation and Cognition Lab at the University of Pennsylvania. Eight participants were excluded due to recording

failures or excessive (more than thirty) mistakes in the task, and the data from the remaining 112 participants is presented here. Of these 112 participants, fifty-six self-identified as female, fifty-three as male, two as nonbinary, and one did not disclose. Since most 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. Due to an error in recording racial/ethnic self-identification on the first half of the surveys administered, we report this data for only sixty-one participants: thirty-seven identified as White, fourteen as Black or African American,[10] four as Asian or Pacific Islander, four as Hispanic or Latinx, and two 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. As in experiment 1, participants were labeled as 'Southern' or 'non-Southern', depending on their residential history. Fifty-two participants were classified as 'Southern', and the remaining sixty participants were classified as 'non-Southern'. In general, the Southern participants produced somewhat more Southern vowels than the non-Southern participants, though the majority of Southern participants were not recognizably Southern-shifted.

3. RESULTS.

3.1. EXPERIMENT 1. Unsurprisingly, the model for experiment 1 shows a significant main effect of Duration (see Table 1), such that longer tokens are produced with a higher front diagonal, indicating a stronger /aɪ/ glide. There is a main effect of Phase, indicating that the baseline-to-exposure shift is significant for those who heard the Southern model talker (reference level). These participants produced a weaker /aɪ/ glide in the exposure (compared to the baseline) phase (est. = −0.135, $p = 0.003$) in the Southern voice condition, indicating a more Southern-like production. There is no effect of Phase for the post-exposure (compared to baseline) phase, suggesting that participants return roughly to their baseline post-exposure (est. = −0.067, $p = 0.135$). The significant interaction between Phase and Voice shows that this baseline-to-exposure shift toward the Southern talker is significantly greater than the (lack of) shift in the Midland control condition (est. = 0.136, $p = 0.038$).

As Figure 2 shows, in the Midland voice condition, participants show virtually no shift across the three experimental phases, whereas in the Southern voice condition, participants shift during the exposure phase to produce weaker glides before returning roughly to their baselines post-exposure. As Figure 3 shows, this effect is driven by Southern participants. Non-Southerners in the Southern voice condition shift minimally in the expected direction, with a slightly lower /aɪ/ glide in the exposure phase, but confidence intervals are largely overlapping. Southerners exhibit larger shifts. The model confirms that Southerners shift more toward the Southern talker than non-Southerners do, evidenced by a significant interaction between Phase(exposure) and Dialect (est. = 0.094, $p = 0.043$).

Unexpectedly, participants randomly assigned to the Midland and Southern voice conditions differ in their baseline productions, evidenced by the main effect of Voice (est. = −0.137, $p = 0.038$) in the baseline phase (reference level). This suggests that more already-Southern-shifted participants were randomly assigned to the Midland condition. Note also that there is no significant main effect of Dialect for the Southern voice (reference-level) condition, suggesting that Southerners and non-Southerners

[10] Since /aɪ/ glide-weakening is also a feature of AAE (both within and outside of the South), it is worth noting that only two of the fourteen known African American participants were non-Southern in the experimental condition.

	EST	*SE*	*df*	*t*-VALUE	*p*-VALUE	
(intercept)	0.104	0.058	150.80	1.794	0.075	.
Phase(exposure)	−0.135	0.045	4458.00	−2.967	0.003	**
Phase(post)	−0.067	0.045	4443.00	−1.496	0.135	
Voice(Midland)	−0.137	0.065	147.80	−2.091	0.038	*
Dialect(non-Southern)	−0.009	0.046	146.50	−0.204	0.839	
Baseline	0.314	0.046	148.70	6.753	< 0.001	***
Duration	0.284	0.023	1088.00	12.354	< 0.001	***
Frequency	−0.069	0.039	58.75	−1.770	0.082	.
Phase(exposure) * Voice(Midland)	0.136	0.065	4445.00	2.079	0.038	*
Phase(post) * Voice(Midland)	0.054	0.065	4441.00	0.837	0.402	
Phase(exposure) * Dialect(non-Southern)	0.094	0.046	4438.00	2.020	0.043	*
Phase(post) * Dialect(non-Southern)	0.032	0.046	4438.00	0.696	0.486	
Voice(Midland) * Dialect(non-Southern)	0.013	0.069	147.50	0.185	0.854	
Phase(exposure) * Baseline	−0.182	0.047	4449.00	−3.917	< 0.001	***
Phase(post) * Baseline	−0.067	0.046	4449.00	−1.442	0.149	
Voice(Midland) * Baseline	0.044	0.069	149.30	0.639	0.524	
Phase(exp) * Voice(Mid) * Dialect(non-S)	0.032	0.069	4437.00	0.462	0.644	
Phase(post) * Voice(Mid) * Dialect(non-S)	0.074	0.069	4441.00	1.082	0.279	
Phase(exp) * Voice(Mid) * Baseline	−0.103	0.069	4447.00	−1.500	0.134	
Phase(post) * Voice(Mid) * Baseline	−0.128	0.069	4451.00	−1.867	0.062	.

TABLE 1. Summary of fixed effects for the linear mixed-effects model predicting the front diagonal (F2 – 2 * F1) at the glide (80%) for experiment 1.

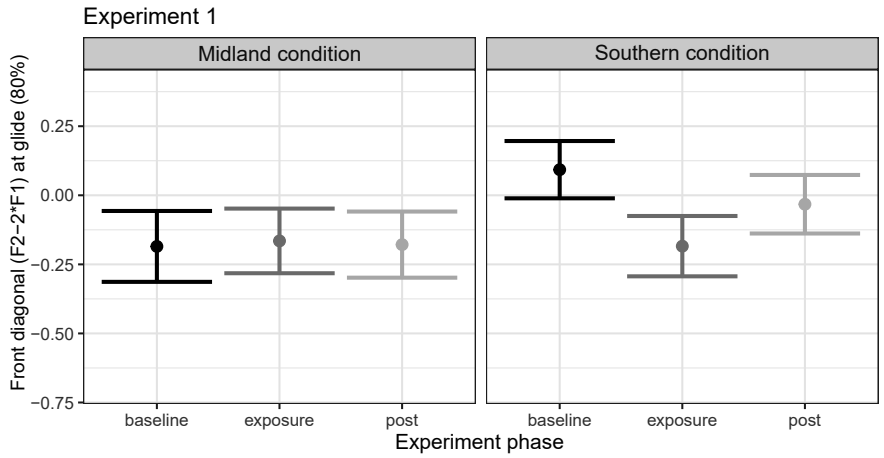

FIGURE 2. Shift across phases by voice condition in experiment 1. Mean values with 95% confidence intervals for all tokens.

have similar baselines in this condition. This means that it is only in the Midland voice (control) condition that the dialect groups differ in their baselines. While this observation does not alter the interpretation of the overall results, it does offer a note of caution that randomly assigned groups taken from the same population may not be as homogeneous as expected. Finally, there is a significant Phase * Baseline interaction for the exposure (compared to baseline) phase (est. = −0.182, $p < 0.001$), suggesting that those with higher and fronter baseline glides shift toward the Southern talker more than those who produced weakened baseline glides. That is, those with /aɪ/ glides most different from what is typical of Southern-shifted speech shift more. This may stem from floor effects (i.e. it is impossible to imitate something you are already doing), greater attention paid to more dissimilar and salient features, or attempts at phonemic contrast main-

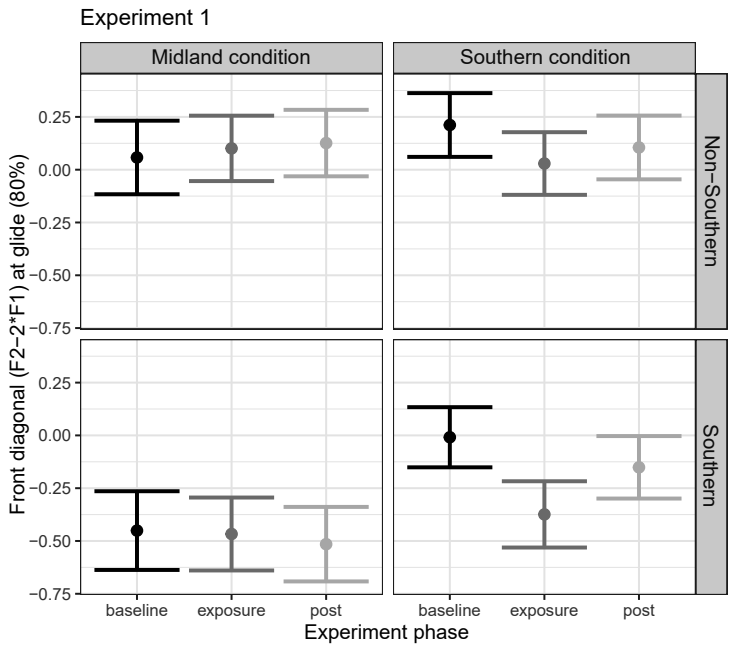

FIGURE 3. Shift across phases by voice condition and participant dialect in experiment 1. Mean values with 95% confidence intervals for all tokens.

tenance; these possibilities are explored in the discussion section (§4). Based on these results, the word-naming game paradigm was deemed suitable for eliciting expectation-driven convergence.

3.2. EXPERIMENT 2. Experiment 2 set out to replicate the results of experiment 1, with several methodological changes intended to increase token counts and test replicability across minimally different conditions. Experiment 2 elicited more target tokens (ninety compared to sixty) from more participants (112 compared to seventy-nine), allowing for more robust statistical analyses. Other changes include new Southern and Midland model talkers, a different subject pool participating online rather than in person, and the inclusion of additional measures examined as individual-differences predictors of convergence, including affective measures and cognitive-style/personality traits. These methodological changes were outlined in §2 above.

EXPECTATION-DRIVEN CONVERGENCE IN THE AGGREGATE. As expected, there are significant main effects of Frequency and Duration (see Table 2), such that shorter tokens and higher-frequency items are produced with a weaker /aɪ/ glide that is lower along the front diagonal. The model also shows a significant main effect of Phase for participants who heard the Southern talker (reference level). These participants shifted from their baselines to produce more Southern-like glides during the exposure phase (est. = −0.129, $p = 0.001$). However, this shift does not last into the post-exposure phase, which does not differ significantly from the baseline phase (est. = −0.044, $p = 0.345$), suggesting that participants return roughly to their baseline productions after exposure.

Figure 4 shows that, 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

	EST	*SE*	*df*	*t*-VALUE	*p*-VALUE	
(intercept)	−0.023	0.057	131.20	−0.399	0.691	
Phase(exposure)	−0.129	0.038	105.76	−3.354	0.001	***
Phase(post)	−0.044	0.046	105.45	−0.949	0.345	
Voice(Midland)	0.050	0.048	81.38	1.054	0.295	
Baseline	0.267	0.042	90.88	6.397	< 0.001	***
Dialect(non-Southern)	0.082	0.035	81.83	2.323	0.023	*
Frequency	−0.103	0.043	87.43	−2.382	0.019	*
Duration	0.522	0.031	176.90	16.690	< 0.001	***
Phase(exposure) * Voice(Midland)	0.147	0.055	103.78	2.675	0.009	**
Phase(post) * Voice(Midland)	0.019	0.066	104.18	0.286	0.776	
Phase(exposure) * Baseline	−0.184	0.048	113.08	−3.841	< 0.001	***
Phase(post) * Baseline	−0.230	0.057	111.90	−4.040	< 0.001	***
Voice(Midland) * Baseline	0.086	0.051	87.90	1.660	0.100	
Phase(exposure) * Dialect(non-Southern)	0.104	0.040	104.80	2.581	0.011	*
Phase(post) * Dialect(non-Southern)	0.098	0.048	103.57	2.033	0.045	*
Voice(Midland) * Dialect(non-Southern)	0.010	0.049	81.66	0.199	0.843	
Phase(exposure) * Voice(Mid) * Baseline	0.026	0.060	112.62	0.432	0.667	
Phase(post) * Voice(Mid) * Baseline	0.014	0.071	111.66	0.192	0.848	
Phase(exp) * Voice(Mid) * Dialect(non-S)	−0.054	0.056	104.03	−0.951	0.344	
Phase(post) * Voice(Mid) * Dialect(non-S)	−0.060	0.068	103.88	−0.895	0.373	

TABLE 2. Summary of fixed effects for the linear mixed-effects model predicting the front diagonal (F2 – 2 * F1) at the glide (80%) for experiment 2.

between Phase and Voice is significant in the model (est. = 0.147, $p = 0.009$), confirming that the baseline-to-exposure shift toward glide-weakened /aɪ/ is greater toward the Southern talker than toward to Midland talker in the 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 (reference-level) productions of /aɪ/, which differs from the unexpected baseline difference between randomly assigned participant groups in experiment 1.

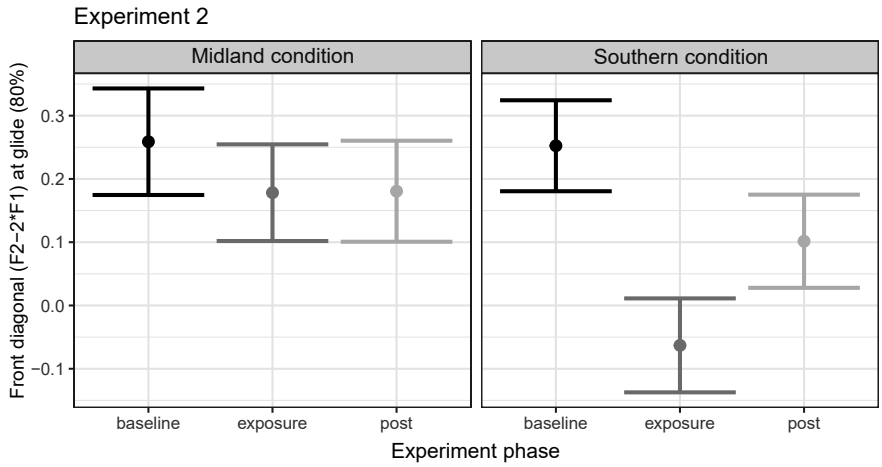

FIGURE 4. Shift across phases by voice condition in experiment 2. Mean values with 95% confidence intervals for all tokens.

The model also shows a main effect of Dialect, such that non-Southerners in the Southern voice condition produce a stronger baseline glide (est. = 0.082, $p = 0.023$) than Southerners, which is the expected dialect difference between the two groups. In

addition to having different baseline productions, Southerners and non-Southerners exhibit different-sized shifts in the Southern voice condition. As the significant interaction between Phase(exposure) and Dialect shows, non-Southerners shift toward the Southern talker less than Southerners do (est. = 0.104, $p = 0.011$). Figure 5 breaks down the aggregate effect shown in Fig. 4 by participant dialect background. While both groups exhibit the same general trend, with a lower glide in the exposure phase, then a shift back toward the baseline in the post-exposure phase, these effects appear to be driven primarily by Southerners, and non-Southern participants exhibit these shifts to a lesser degree. All three phases show overlapping confidence intervals, and mean shifts are in general smaller for non-Southerners. Also note that Southerners do not return fully to their baselines in the post-exposure phase as non-Southerners do, evidenced by a significant interaction between Phase(post) and Dialect in the Southern voice condition (est. = 0.098, $p = 0.045$), though this could be because non-Southerners shift less to begin with. Regardless, Southerners appear to exhibit some residual convergence toward the Southern talker, even after exposure.

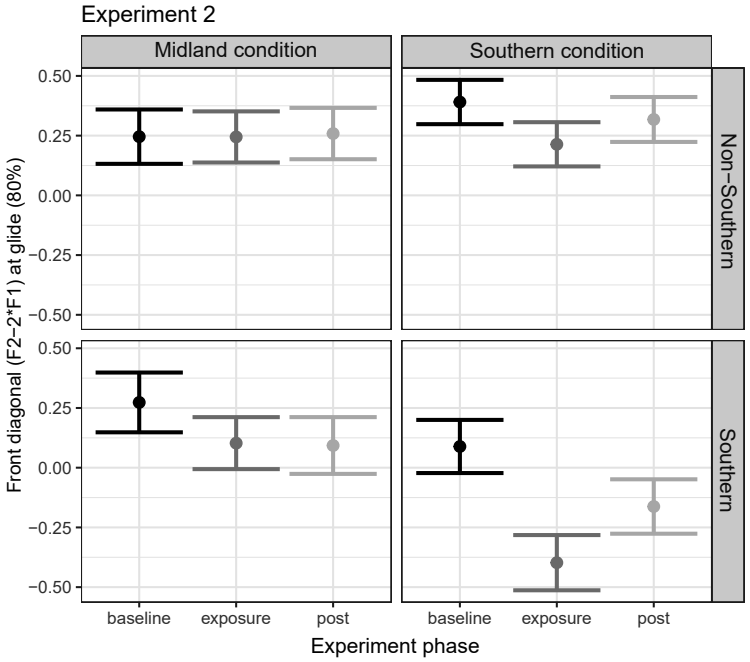

FIGURE 5. Shift across phases by voice condition and participant dialect in experiment 2. Mean values with 95% confidence intervals for all tokens.

Finally, there is a role for participant baseline in degree of shift, evidenced by a significant interaction between Phase(exposure) and Baseline (est. = −0.184, $p < 0.001$). Those with higher and fronter /aɪ/ glides to begin with shift more toward the Southern model talker than those with already weaker glides. There is also a significant interaction between Phase and Baseline for the post-exposure (compared to baseline) phase (est. = −0.230, $p < 0.001$), suggesting that those with higher/fronter baseline glides not only shift more but also stay shifted longer into the post-exposure phase. Note that the effects of Dialect and Baseline on degree of shift toward the Southern voice are in opposite directions. That is, Southerners (who tend to have weaker glides to begin with)

shift more than non-Southerners, yet those with already weaker glides shift less. While these two observations may appear to be at odds with one another, it may in fact indicate that speakers shift toward the most salient features distinct from their own productions, as long as they have sufficient experience and their repertoires allow it. These possibilities are explored in more detail in §4.

We may wonder at this point whether non-Southerners shift less than Southerners because they are shifting in other dimensions not captured in the realization of the glide (F2 – 2 * F1 at 80% measure). Since non-Southerners are less familiar with the Southern dialect, attempts at converging toward glide-weakened /aɪ/ may not be apparent at the 80% mark alone. Formant trajectories for both F1 and F2 are visualized in Figure 6. 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 six time points for all tokens.

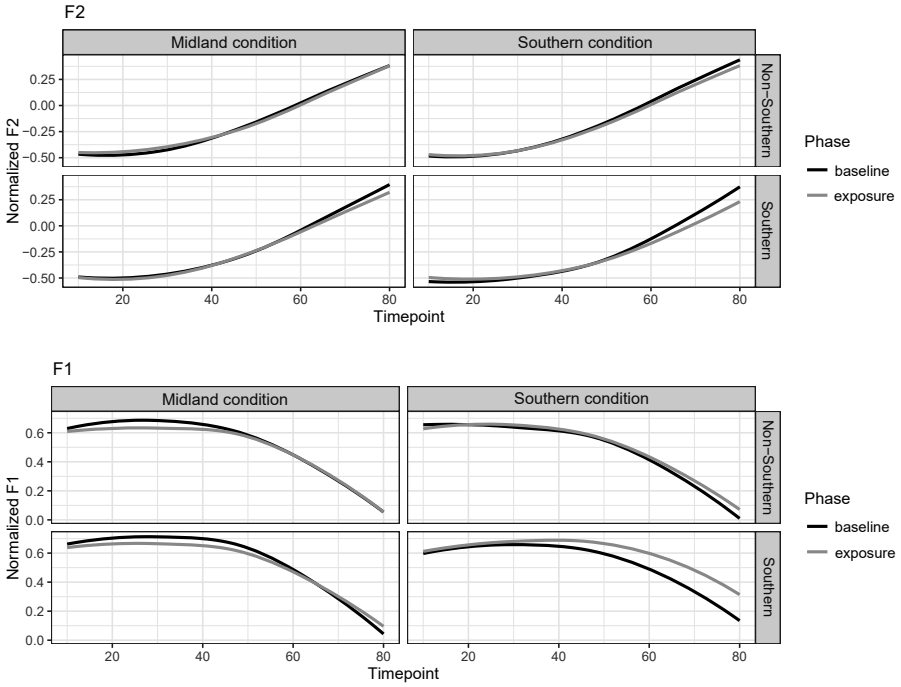

FIGURE 6. Formant trajectories for F1 (bottom) and F2 (top), experiment 2.

As we can see, there are minimal shifts in F2 at the glide across phases, with Southerners shifting toward lower F2 values throughout the course of the experiment in the Southern voice condition, consistent with a more Southern production. Note, however, that similar (smaller) shifts in this same direction occur for Southerners in the Midland condition, suggesting that this shift may be partially due to global shifts in speech as the experiment progresses. Non-Southerners exhibit minimal shifts in the same direction, but only in the Southern voice condition. In F1, there is a general shift for both dialect groups in the Midland control condition, such that there is a lower F1 at the nucleus, consistent with /aɪ/ nucleus raising. It is unclear why both groups would respond to the Midland talker (but not the Southern talker) in this way, though it is possible that they

are positing a raised /aɪ/ vowel for the Midland talker and converging toward this target. The clearest baseline-to-exposure difference in Fig. 6 comes from Southerners in the Southern voice condition, who shift to produce a higher F1 in the latter half of the vowel, with the greatest shift at the 80% mark. Non-Southerners show similar (but much smaller) shifts, consistent with what we see in Fig. 5. We can conclude then that participants in the aggregate do not appear to be shifting toward (or away from) the Southern talker in some unexpected way not captured by our measure of the phonetic realization of the glide at the 80% mark.

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 7 shows, nearly a third (32%) of non-Southerners shift in the opposite direction (i.e. they diverge from the Southern model talker) but only five (or 18.5%) of Southern participants diverge. The remainder of this section investigates various individual traits that may facilitate individual convergence to the Southern model talker and finds little evidence for individual-differences predictors of expectation-driven convergence. To investigate the role of various individual-differences measures in convergence, I correlate each measure to an estimate of individual shift: individual random effects were extracted from the model presented in Table 2 for the exposure (compared to baseline) phase using the coef() function in R. This is referred to as the 'baseline-to-exposure shift' in the following analyses, and a smaller value (indicating a more weakened glide in the exposure phase) indicates greater convergence, while a larger value indicates less convergence.

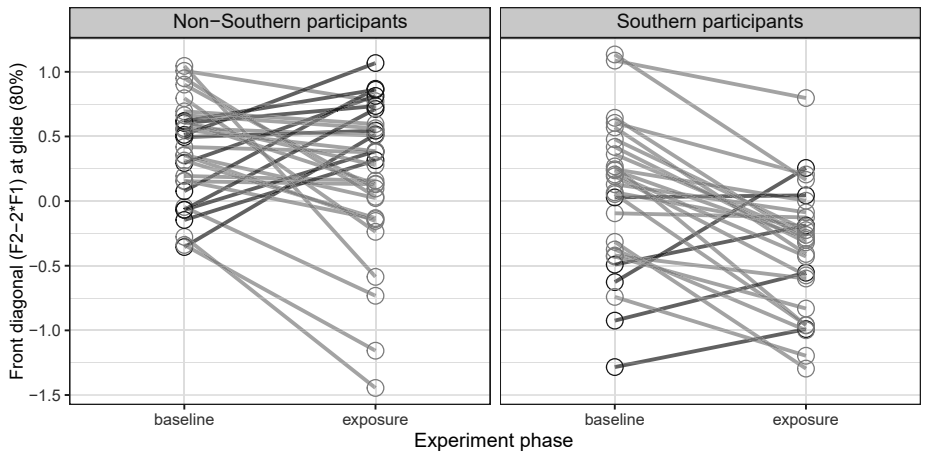

Figure 7. Individual differences for those in the Southern voice condition in experiment 2. Gray indicates convergence. Black indicates divergence.

It is important to correct *p*-values for multiple comparisons since several correlations are being conducted on the same data set. This is due to the fact that the greater the number of correlations tested, the greater the likelihood of finding false positive results. All of the correlations presented here are Pearson's correlations that have been adjusted with permutation tests using the perm.relation() function in the wPerm package in R (Weiss 2015). A permutation test is a means of multiple-comparisons correction, where

one variable is randomized (i.e. the same values are randomly assigned to different participants) such that the relationship between the two variables being correlated is broken but the structure of the data set and distribution of the two variables remain the same. Each randomization of the data represents a possible sampling of individuals under the null hypothesis (i.e. there is no correlation between the two variables). This randomization is repeated a set number of times (here, $N = 10{,}000$), and the p-value resulting from these tests indicates the likelihood of getting the same, or greater, Pearson's R under a random distribution (i.e. under the null hypothesis).

None of our affective or familiarity measures predicted degree of convergence shift, including Prestige and Likability of the model talker and Familiarity with the South, and there is likewise little evidence that desire for positive social evaluation, as measured by the Marlowe-Crowne social desirability scale, facilitates convergence. The lack of correlation for these features is shown in Appendix B. Further, no demographic features tested (age, gender, race, education) significantly predicted degree of convergence.

Of the multiple individual-differences measures investigated, only four correlate with degree of convergence, though this differs based on dialect group. The autism-spectrum quotient (AQ; Baron-Cohen et al. 2001) subscores ATTENTION TO DETAIL (AD) and IMAGINATION (I) both correlate with degree of shift, but only for non-Southerners, such that more 'autistic-like' traits on these scales result in LESS convergence (AD: Pearson's $R = 0.502$, $p = 0.006$; I: Pearson's $R = 0.413$, $p = 0.026$). While neither subscore of the AQ has previously been found to correlate with degree of convergence shift, there is some evidence that high attention measures (such as high ATTENTION-SWITCHING (AS) scores on the AQ) lead to increased convergence. However, we actually observe the opposite pattern here, with higher AD scores (indicating greater attention to detail) correlating with higher individual coefficients for the exposure phase (indicating a higher and fronter /aɪ/ glide and therefore less convergence); see Figure 8. Note, however, that we would not necessarily predict greater attention to phonetic detail to increase the amount of EXPECTATION-DRIVEN convergence (even though it has been observed to correlate with degree of INPUT-DRIVEN convergence), because there is no local phonetic signal to direct increased attention toward. Note additionally that shifts toward weaker /aɪ/ glides also unexpectedly correlate with I scores in the same direction for non-Southerners in the MIDLAND voice condition ($R = 0.452$, $p = 0.02$), so it appears that the role of the I subscore actually has more to do with general shift (perhaps as a reduction process due to fatigue) than with convergence toward the Southern talker.

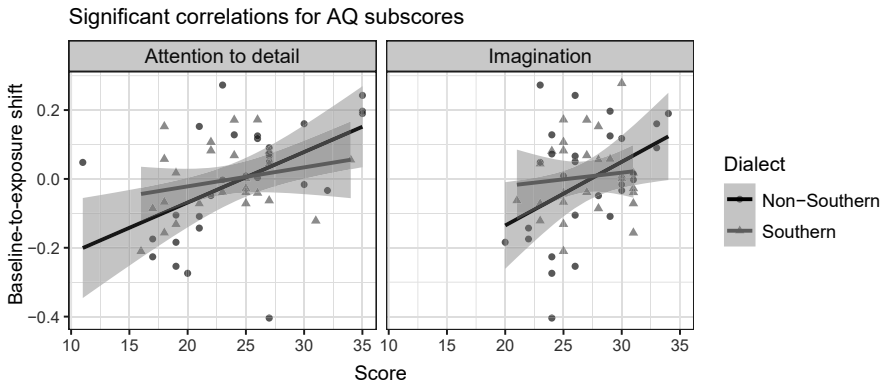

FIGURE 8. Significant correlations for AQ subscores and degree of convergence shift in the Southern voice condition in experiment 2. Higher shift values on the y-axis indicate a higher/fronter glide in the exposure phase, while lower values indicate a weaker glide in the exposure phase (i.e. convergence).

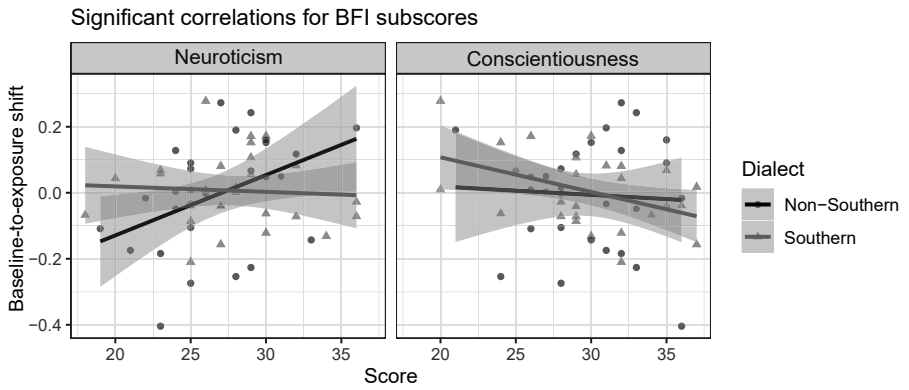

FIGURE 9. Significant correlations for BFI subscores and degree of convergence shift in the Southern voice condition. Higher shift values on the *y*-axis indicate a higher/fronter glide in the exposure phase, while lower values indicate a weaker glide in the exposure phase (i.e. convergence).

Finally, as seen in Figure 9, two Big Five inventory (BFI; John et al. 1991) measures correlated with degree of convergence: lower neuroticism scores indicated greater convergence (contradicting previous findings of e.g. Lewandowski and Jilka (2019)), but only for non-Southerners ($R = 0.414$, $p = 0.029$), and higher conscientiousness scores indicated greater convergence, but only for Southerners ($R = 0.435$, $p = 0.023$). However, conscientiousness scores also correlated in the same direction with shifts in the Midland control condition for Southerners ($R = -0.513$, $p = 0.007$), again suggesting that this is a predictor of general shift rather than convergence. No measure seemed to predict convergence for both groups, two of the measures flat-out contradict previous findings on the role of individual-differences predictors in degree of convergence shift, and two of the measures correlate with shifts in both the experimental and control conditions.[11] There is therefore little evidence that individual differences as measured here are predictive of expectation-driven convergence behavior.

4. DISCUSSION. The results of the word-naming game task have provided evidence that speakers do shift their speech toward expected linguistic behavior. Participants in the aggregate produced more Southern-like, glide-weakened tokens of the /aɪ/ vowel after hearing a Southern talker, even though the talker produced no instances of /aɪ/. Those in the control condition who heard a Midland talker did not exhibit such shifts, suggesting 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 or acclimation to the task) as the experiment progressed. Findings also suggest that the word-naming game paradigm is suitable for eliciting convergence experimentally and may prove a useful method for tackling other questions in linguistics. Further, the results of experiment 2 precisely mirror those of experiment 1. Showing that the main findings can be replicated with a different model talker and subject pool suggests that expectation-driven convergence is a real and reliable linguistic behavior. These findings support anecdotal accounts of the existence of expectation-driven convergence (Bell 1984, Auer & Hinskens 2005) and also align nicely with the findings of Wade and Roberts (2020), who found experimental evidence of expectation-driven convergence in computer-mediated communication in an artificial 'alien' language.

[11] Note that none of these significant measures correlated significantly with participants' baselines.

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, as participants in the aggregate begin to shift back to their baselines after the exposure phase. Since participants returned to reading clues rather than hearing them post-exposure, with their pseudo-interlocutor no longer present, they may have simply lacked the social motivation to converge, which may provide some evidence that expectation-driven convergence differs from other socially triggered production shifts in that it persists only while an interlocutor is present. Alternatively, in the absence of a trigger in the post-exposure phase, participants may have no longer been cognitively primed to converge. Regardless, the existence of expectation-driven convergence supports a growing body of evidence (see Foulkes & Docherty 2006, Thomas 2011, and Campbell-Kibler 2016 for overviews) that phonetic detail is stored in long-term memory and is in some way linked to social knowledge, and that speakers utilize this knowledge in speech perception and production.

4.1. Individual differences. The experiments presented here have demonstrated that participants exhibit considerable individual differences in production behavior. Most notably, not all participants in the Southern voice condition converged as expected. One source of these differences is exposure to the Southern dialect—only 18.5% of Southern participants failed to shift toward the Southern talker, while nearly one third of non-Southern participants failed to converge, and Southerners in the aggregate converged significantly more than non-Southerners. This is not surprising, since those familiar with Southern speech should be better able to replicate the phonetic details of /aɪ/ glide-weakening based on expectations alone; it is likely more important for the forms a speaker converges toward to be familiar to them when there are no instances of these forms available as models within the context of the experiment. As Bell (1984:190) notes regarding 'outgroup design': '[t]he 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'. However, self-reported familiarity with the South did not predict degree of shift toward the Southern talker for either dialect group. While it is possible that the 'familiarity' measure used here was simply not sensitive to differences in degree of exposure to glide-weakened /aɪ/, an alternative reason why Southerners may have converged more may have less to do with familiarity and more to do with the fact that glide-weakened /aɪ/ was already within their repertoire (i.e. they had experience producing it).

While Southern participants converged more than non-Southerners, it is also the case that the least Southern-sounding participants in either dialect group tended to converge the most. One simple possibility for the role of baseline production in degree of convergence is a floor effect: it is impossible for an individual to converge toward something they are already doing. Even if an individual does align with a Southern talker by precisely imitating their expected /aɪ/ target, this will not appear empirically as a production shift if the participant was producing glide-weakened /aɪ/ to begin with. However, floor effects cannot fully account for the fact that non-Southerners, who did not produce already glide-weakened /aɪ/ prior to exposure, also exhibited a similar baseline effect. A more plausible explanation may have to do with variant salience. It has often been observed that convergence occurs to salient features but not less salient

features. For instance, Zellou et al. (2013) found that reduced nasality was imitated for longer than increased nasality, which they proposed resulted from the greater novelty and perceptual salience of reduced nasality. Similarly, Podlipský and Šimáčková (2015) argued that imitation of extended (but not reduced) prevoicing in Czech was a result of the greater salience of extended prevoicing. Salience seems to be a likely explanation here, since variants more acoustically dissimilar to one's own are more likely to be noticed in the first place.

Another explanation for divergence/maintenance among participants with weaker baseline glides could be attempts at contrast preservation. 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/. Contrast preservation has previously been proposed as exerting potentially inhibitive influences on convergence processes. 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 prevoicing was not imitated and suggested that this was because it was not relevant for phonological contrasts (though noncontrastive features may also be less salient, which could account for these findings). While the baseline effect may at first seem contradictory to the finding that more Southerners than non-Southerners converged, it actually aligns nicely with previous studies such as Walker & Campbell-Kibler 2015, which found that the most phonetically distinct tokens (those that are novel and therefore most salient) elicit greater convergence, but only if the target is within a speaker's repertoire (and as long as they were unaware of it).

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 such 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. Alternatively, attitudes toward the talker may not have been as relevant as other factors in determining whether (or how much) convergence occurred. For instance, a non-Southern participant may have very positive feelings toward the Southern model talker, but without sufficient experience with Southern speech, they may not be able to converge. Additionally, the role of attitude in convergence is generally quite small, and the experiment simply may not have picked up on such a small effect, particularly if other factors were simultaneously impacting degree of convergence.

Likewise, there was little evidence that cognitive-style/personality traits as measured by the BFI, AQ, and Marlowe-Crowne social desirability scale play a role in convergence behaviors. In fact, the present study found that neuroticism scores on the BFI predicted LESS convergence for non-Southerners, contradicting Lewandowski and Jilka (2019), who found that neuroticism scores (which they interpret as signaling a higher need for social approval) and lower BEHAVIOR INHIBITION SCALE (BIS) scores (which measure motivation to avoid aversive outcomes) predicted GREATER convergence in a cooperative Diapix task. Relatedly, the AD subscore of the AQ also predicted LESS convergence for non-Southerners, contradicting previous reports that increased attention,

which may affect the level of engagement with exposure materials and greater attention paid to fine phonetic details, facilitates phonetic convergence. Yu et al. (2013), for instance, found that individuals with greater openness and higher AS scores on the BFI and AQ 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.

However, it is not so surprising that attentional measures do not predict increased expectation-driven convergence; because there is no phonetic signal for /aɪ/ to which participants can attune in the course of the experiment, greater attention to phonetic detail (at least locally) would not be useful in targeting nonobserved glide-weakened /aɪ/. The finding that greater attention to detail actually predicts LESS convergence could indicate that, when those with greater attention to detail fail to observe a local target for a participant's /aɪ/ glide, they are more likely to maintain their own speech norms—though this is quite speculative. This would not explain why higher neuroticism scores also indicated less convergence, since greater desire for social approval would similarly be predicted to facilitate expectation-driven convergence (perhaps more so than for input-driven convergence, because it is less straightforwardly understood as a potentially automatic consequence of perception). Finally, both the imagination subscore on the AQ and the conscientiousness measure of the BFI correlated with convergence, but these factors also correlated with shifts toward glide-weakened /aɪ/ in the Midland (control) condition, suggesting that they were really picking up on a tendency to produce more centralized vowel targets over time (i.e. due to fatigue or acclimation to the task) rather than convergence behaviors. While individual-differences measures do not offer much insight into an individual's tendency to exhibit expectation-driven convergence, these results do suggest that caution should be exercised in interpreting correlations between individual-differences measures and production shifts, and they indicate the need for more work on the replicability of previous findings across different types of convergence behaviors, such as with expectation-driven convergence specifically.

4.2. THE MECHANISMS DRIVING CONVERGENCE. This new evidence for the existence of expectation-driven convergence has broader implications, including for theories of the mechanisms behind convergence. Automatic accounts of linguistic convergence typically assume tight perception-production linkages, such that convergence is an automatic consequence of perception. Under episodic models of speech perception and production (e.g. Goldinger 1998, Goldinger & Azuma 2004), lexical items are stored in memory with detailed phonetic information attached to each instance. '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 and Goldinger & Azuma 2004, imitation occurred only for identical lexical items (though note that later studies have not replicated lexical specificity in convergence), providing support for an exemplar model that is lexically specific. Imitation was also found to be greater for lower-frequency words (Goldinger

1998), which supports an episodic usage-based model in that the fewer similarly activated traces there are to compete with the prime words listeners heard in the experiment, the more influential the memory traces of these prime words would be. Accounts of convergence appealing to structural priming mechanisms also assume tight perception-production relationships and are considered a route for automatic perceptual influences on production. The INTERACTIVE ALIGNMENT MODEL (Pickering & Garrod 2004), for instance, suggests that individuals align their linguistic representations due to automatic priming mechanisms that serve to increase intelligibility for both speakers. Under a priming account, a recently heard form spreads its activation to similar or related forms, facilitating their access and promoting their usage.

Other accounts under which perception and production are tightly linked include gestural-based accounts of speech perception, which predict convergence on the basis of gestural perception. In MOTOR THEORY (Liberman & Mattingly 1985), speech perception units consist of articulatory gestures rather than acoustic properties, such that gestures are automatically 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 & Garrod 2013, Gambi & Pickering 2013). Under this account, 'actors construct forward models of their actions before they execute those actions, and … perceivers of others' actions covertly imitate those actions, then construct forward models of those actions' (Pickering & Garrod 2013:329). Support for gestural accounts of convergence comes from evidence that phonetic imitation occurs when model utterances are presented only visually (Gentilucci & Bernardis 2007, Miller et al. 2010).

Automatic accounts such as those described above tend to focus on input-driven rather than expectation-driven convergence, since they rely on maximally similar perception triggers and production targets that share a linguistic representation at some level. Under direct realism and motor theory, for instance, imitation is a natural consequence of employing the motor system in linguistic perception, but a recently perceived motor gesture would have to be similar enough to the production target to exert any real influence. (In other words, perceiving Southern-shifted /e/, for example, should not impact subsequent productions of a different vowel like /aɪ/.) Similarly, episodic models may account for convergence via automatic activation of exemplars, though such accounts often require a shared lexical representation. Under some models, such as that in Goldinger 1998, exemplars consist of lexical entries, meaning that imitation would occur only toward the same lexical item recently heard. Even exemplar models where perception triggers and production targets come from different exemplar pools consisting of different-sized segments (i.e. not just lexical items) do not straightforwardly account for expectation-driven convergence with the typically proposed explanations of how and why imitation occurs. Under such a dual-representation model (Garrett & Johnson 2013), speech-perception exemplars are broader and involve lexical segments, while production exemplars are narrower (accounting for the fact that we can successfully perceive greater variation than we can produce) and stored in smaller segment-sized units. Regardless, such a model still proposes that production targets are derived from the average of activated targets in the production exemplar pool, and crucially 'among the many factors that determine exemplar activation THE INTENDED LINGUISTIC CATEGORY obviously matters a great deal'

(Garrett & Johnson 2013:90, emphasis mine). As such, imitation would not be facilitated by activation of exemplars that are not of the same linguistic category.

Expectation-driven convergence instead requires utilization of mental associations between representationally distinct things; therefore, we are not able to explain expectation-driven convergence with many of the automatic accounts previously proposed for input-driven convergence, which rely on perception-turned-production mechanisms. This is not to say that input-driven convergence is not accomplished via any of the mechanisms described above, nor that the mechanisms behind expectation-driven convergence cannot involve automatic perception-production links. Rather, this simply means that convergence is capable of targeting non-locally observed forms. In fact, expectation-driven convergence is still compatible with an automatic perception-production relationship that relies on associative links. Hybrid implementations of exemplar models, and other models of linguistic processing that allow for encoding of social and contextual information alongside detailed phonetic information (Foulkes & Docherty 2006, Johnson 2006), are actually well suited to explain how sociolinguistic associations are formed and utilized. Indeed, exemplar-theoretic models may leave room for the influence of activated social categories on associated linguistic representations—but these are seldom employed as explanations for imitation specifically. Under such models, fine phonetic detail is stored alongside other information about an utterance, such as details of the social characteristics of whoever spoke the utterance and contextual information like utterances spoken alongside it.

Expectation-driven convergence may then be governed by the same sort of associative priming mechanisms proposed for input-driven convergence, but may differ simply in that these mechanisms are triggered by different types of information. For example, input-driven convergence toward glide-weakened /aɪ/ is assumed to occur due to its activation via direct and immediate perception of glide-weakened /aɪ/; EXPECTATION-DRIVEN convergence may also be an automatic consequence of activating representations of glide-weakened /aɪ/, but via a different associative route. In the present study, at least two different routes are possible: (i) perceiving other Southern-shifted vowels frequently encountered alongside glide-weakened /aɪ/ may in turn activate representations of glide-weakened /aɪ/ due to their typical cooccurrence in Southerners' speech, or (ii) hearing a Southern model talker may activate the relevant social concept, such as 'Southern', which could in turn activate representations of glide-weakened /aɪ/. It is appealing to account for both input-driven and expectation-driven convergence with the same mechanism—one that involves activation of glide-weakened /aɪ/ via SOME sort of trigger. It also highlights the fact that, in instances of naturalistic spontaneous imitation that is often assumed to be input-driven, we do not actually know how much can be attributed to expectations and activation of social categories. Regardless, focusing on the aspect of exemplar models that allows for multiple associative links across representationally distinct units may prove fruitful in accounting for both expectation-driven and input-driven convergence simultaneously, as well as other behaviors that tap into sociolinguistic knowledge.

It is worth noting that, even within a primarily automatic account of convergence, there is still room for socially motivated processes. One possibility is that automatic influences on production are actively inhibited or otherwise modulated by social factors such as desire for affiliation or to maintain social distance. While ACTIVATION of glide-weakened /aɪ/ may be an automatic consequence of perceiving associated cues, the resulting production shifts may require some degree of agency, 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.

A second possibility is that social motivation may influence perception and production behavior WITHOUT CONSCIOUS CONTROL. Though it has been a commonly held view that social motivation requires awareness and that behaviors we are not aware—or in control—of therefore cannot be social, recent work 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, summarizes recent evidence that many social cognition behaviors observed must occur very rapidly and without conscious awareness. As Campbell-Kibler 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 manipulated intentions' (2016:134, 135). In fact, work on social cognition has found an influence of social information on linguistic processing in as little as 200–300 ms (Van Berkum et al. 2008).

Of course, none of the measures of social motivation for convergence included here correlated with degree of expectation-driven production shifts, so the role of social factors in expectation-driven convergence is an open question, though it is related to the question of social motivation in production shifts more generally. Rather than asking whether automatic accounts or those that allow for the integration of social information are better suited to account for either type of linguistic convergence, I would argue that a more appropriate question to ask is HOW linguistic and social information and perception-production links work in tandem, particularly in the case of expectation-driven convergence. This line of inquiry may also prove useful in reconciling mechanistic and identity-based accounts of convergence. There are a number of ways that automaticity and social motivation may interact in convergence, and in perception and production more generally, and the influence of social factors on the activation, retrieval, and utilization of sociolinguistic knowledge is a future direction worth exploring.

4.3. IMPLICATIONS FOR SOCIOLINGUISTIC COGNITION. An ideal model of sociolinguistic cognition should be able to account for what we have observed in experiments 1 and 2—that perceiving a variant (or a cluster of variants) can lead to activation of a different variant, whether indirectly through activation of a shared social meaning or through more direct links across commonly cooccurring variants. Others have also observed that exposure to one feature facilitates perception or production of related features. Campbell-Kibler (2012) found that, in an IMPLICIT ASSOCIATION TASK, the *-in* realization of the (ING) variable (as in *talking* vs. *talkin'*) was indexically linked with /aɪ/ monophthongization, as both are associated with Southern speech. Vaughn and Kendall (2019) found that when speakers were instructed to produce sentences with only the *-in* variant, they additionally shifted to producing more monophthongal /aɪ/, again likely due to their similar association with Southernness.

The precise associative route through which these types of behaviors are made possible is still unclear. One possibility is that activation of a related variant occurs INDIRECTLY via a shared social category (i.e. 'Southern'). This explanation relies on both the observed variant(s) and the unobserved variant (in our case /aɪ/) indexing a shared social meaning.[12] Another possibility is that producing or perceiving variants can more directly activate other variants that commonly cooccur within individual speakers with-

[12] I have been referring to this here as 'Southern', but in practice it could be some other social category or descriptor such as 'rural'. The social category is not crucial to the indirect associative account except in that it is shared by both variants

out necessarily activating the social category that the variants or speakers share. Kim and Drager (2017), for instance, suggest that cues to age within a phonetic signal can prime associated lexical items DIRECTLY (i.e. without accessing mental representations of the social category encompassing these forms) because they tend to be produced by the same speakers. This direct route may be more likely for expectation-driven behaviors where a shared social meaning is not apparent (e.g. Theodore et al.'s (2009) finding that observed VOT for a talker's production of /p/ informed listener predictions about the same talker's VOT for /k/). There is some evidence that both direct and indirect routes may be utilized, but that it may vary based on dialect experience: in a word-naming game paradigm similar to the task used here (Wade 2020), participants listened to either a Midland talker who was labeled as 'Southern' or a Southern talker labeled as 'Midland'. Participants from the South relied on bottom-up acoustic cues, converging more toward the Southern talker (despite the Midland label), while participants from outside of the South relied more on top-down social category labels and converged more toward the Southern-labeled talker (despite the Midland accent), reflecting the role of experience in the encoding of sociolinguistic information. A model of sociolinguistic cognition in which features are associatively linked with both cooccurring features and social categories may account for both types of routes in the activation of the same sociolinguistic knowledge.

Regardless of the route through which it occurs, the existence of expectation-driven convergence supports an account of linguistic representation in which units larger than individual variants—such as styles or individual systems—are mentally represented. Probing the cognitive reality of larger units such as 'sociolects' is a driving force behind a recent surge of research on the covariation of multiple socially linked variants, or SOCIOLECTAL COHERENCE (Guy 2013, Becker 2016, Guy & Hinskens 2016, Tamminga 2019). Expectation-driven convergence and other findings indicating that larger units may be mentally represented provide an additional source of evidence for the cognitive reality of sociolects. Other recent studies have also provided evidence that larger units may be mentally represented, including that mentioned above by Vaughn and Kendall (2019), in which participants additionally produced monophthongal /aɪ/ and other vowel realizations consistent with a general shift toward a 'Southern' speech style when instructed to produce sentences with *-in*. Present findings also align with those of Zellou et al. (2017), who found that participants shifted toward a model talker's average nasalization throughout the course of an experiment rather than toward the most recent individual tokens.

The present study also contributes to a growing body of research showing that macrosocial categories like dialect region are mentally represented and linked to linguistic features (e.g. Hay & Drager 2010, Campbell-Kibler 2012, Vaughn & Kendall 2019). While microsocial categories like personae have also been shown to be mentally represented and linked to linguistic representations (e.g. D'Onofrio 2015), we know little about the limit of such mental associations. For instance, Rácz et al. (2017, 2020) show that age, ethnicity, and gender (but not the spatial orientation of an interlocutor) are learned in association with linguistic variation, which they suggest is due to the frequency with which age, ethnicity, and gender (as opposed to spatial orientation) have been indexically linked with linguistic features in listeners' past experience.

Likewise, it remains unknown how socially salient a linguistic feature must be in order to exhibit such associative relationships with social categories. The present study focuses on the socially salient glide-weakened /aɪ/, but it is unclear whether similar re-

sults would be found for variants that show similar social distributions in the world but are not as socially salient. It is possible that the 'stereotype' status of /aɪ/ glide-weakening is what promoted expectation-driven convergence from both Southerners and non-Southerners. The lack of effect of familiarity on convergence might suggest that experience with the Southern dialect is not as important in promoting shifts toward glide-weakened /aɪ/ as might have been predicted under exemplar models, where more frequent encounters with a variant should yield more activated glide-weakened exemplars and therefore larger shifts toward glide weakening. This echoes the findings of Walker (2019), who found that US expatriates to the UK and US nonmigrants both exhibited similar-sized topic-based shifts toward using aspirated or glottal /t/ even though they had different amounts of exposure to this feature. Similarly, Sanchez et al. (2015) found that New Zealanders shifted to producing more Australian-like variants when primed with the concept of Australia, 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. Exploring whether there are limits to the types of social categories and linguistic features that can be representationally linked (including the role of salience) is likely a promising area for future research.

5. CONCLUSION. In summary, this article offers new experimental evidence for the existence of expectation-driven convergence and develops a novel methodology for investigating this phenomenon that may also prove useful for other work exploring sociolinguistic expectations. These findings shed new light on questions about the mechanisms underlying linguistic convergence, including the relative automaticity of convergence behaviors and the role of social factors in convergence. Findings also contribute to a growing body of research in sociolinguistic cognition showing that social and linguistic knowledge are linked in the minds of speakers, and that this knowledge is utilized not only in speech perception but also in speech production.

APPENDIX A: MODEL-TALKER VOWEL PLOTS

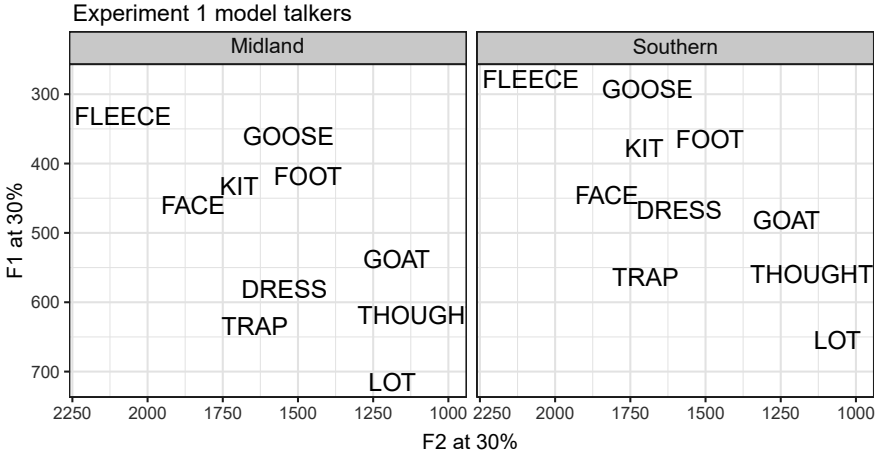

FIGURE A1. Vowel spaces of model talkers for experiment 1, based on their production of all clues and instructions in the experiment.

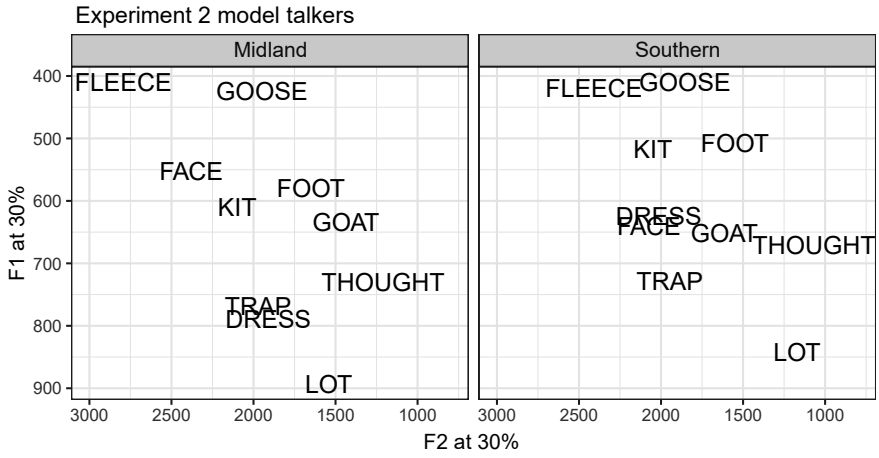

FIGURE A2. Vowel spaces of model talkers for experiment 2, based on their production of all clues and instructions in the experiment.

APPENDIX B: INDIVIDUAL-DIFFERENCES CORRELATIONS

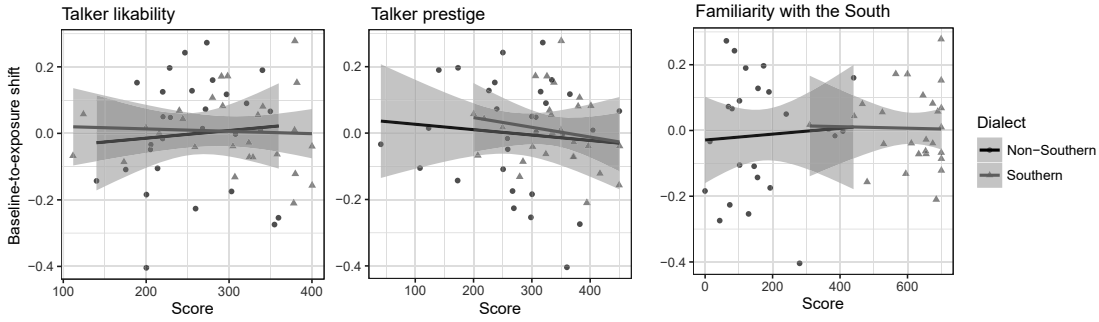

FIGURE A3. Correlation between attitudinal measures and degree of convergence shift, by dialect background. Graphs show values only for those in the Southern voice condition.

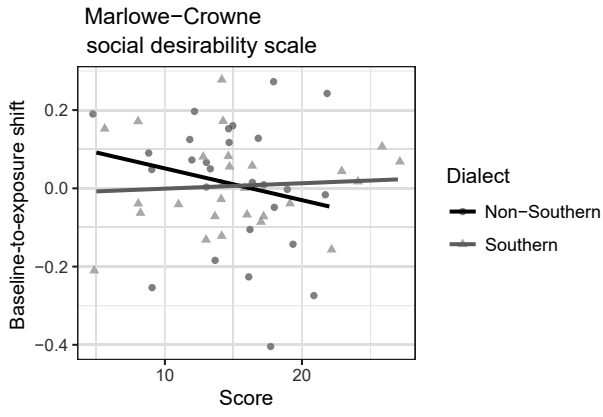

FIGURE A4. Correlation between Marlowe-Crowne social desirability scale scores and degree of convergence shift, by dialect background. Graph shows values only for those in the Southern voice condition.

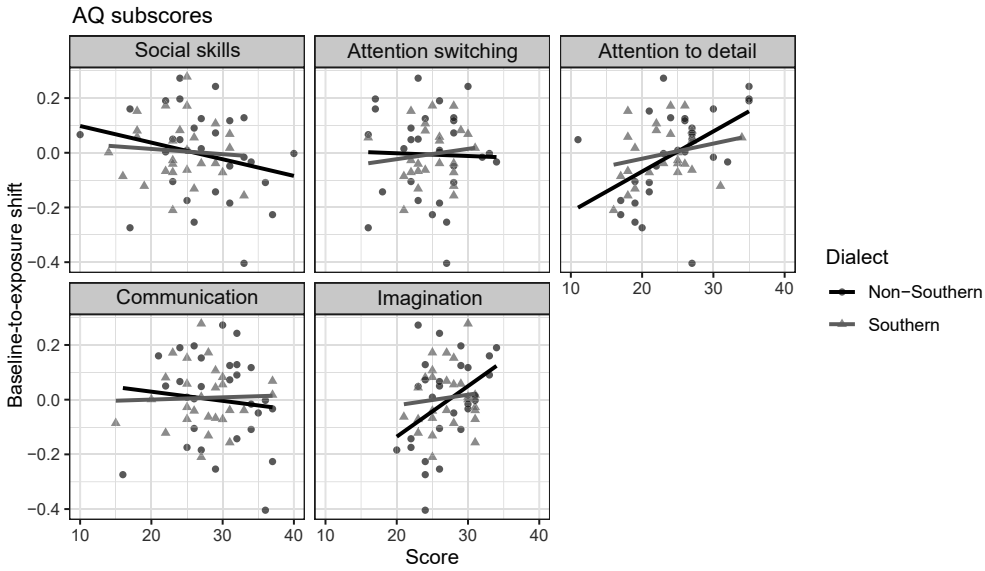

FIGURE A5. Correlation between AQ subscores and degree of convergence shift, by dialect background. Graphs show values only for those in the Southern voice condition.

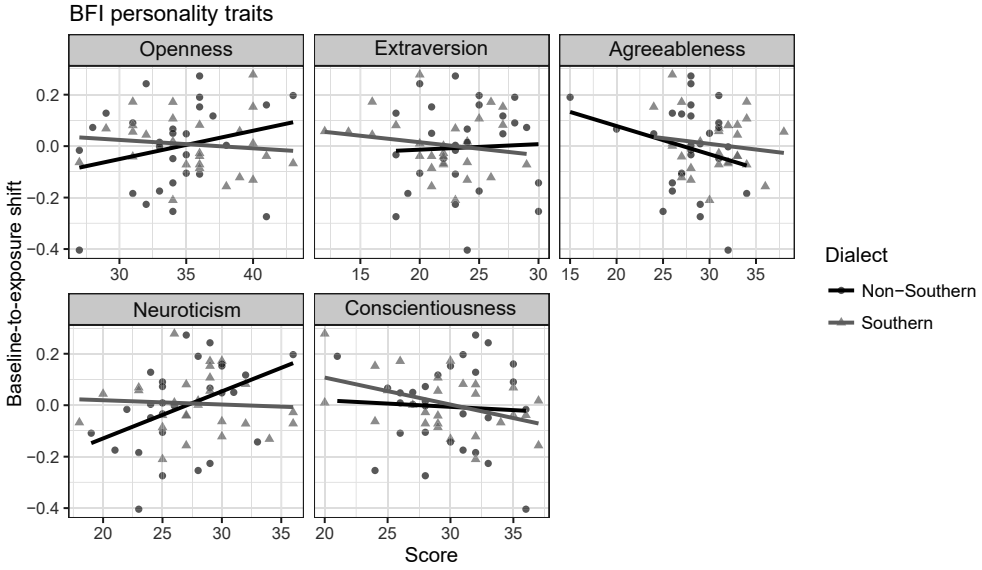

FIGURE A6. Correlation between BFI personality dimensions and degree of convergence shift, by dialect background. Graphs show values only for those in the Southern voice condition.

REFERENCES

AUER, PETER, and FRANS HINSKENS. 2005. The role of interpersonal accommodation in a theory of language change. *Dialect change: Convergence and divergence in European languages*, ed. by Peter Auer, Frans Hinskens, and Paul Kerswill, 335–57. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486623.015.

Bailey, Guy. 1997. When did Southern American English begin? *Englishes around the world: Studies in honor of Manfred Gorlach. Vol. 1: General studies, British Isles, North America*, ed. by Edgar W. Schneider, 255–75. Amsterdam: John Benjamins.

Baron-Cohen, Simon; Sally Wheelwright; Richard Skinner; Joanne Martin; and Emma Clubley. 2001. The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders* 31.5–17. DOI: 10.1023/A:1005653411471.

Becker, Kara. 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. DOI: 10.1016/j.lingua.2015.10.017.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2).145–204. DOI: 10.1017/S004740450001037X.

Bell, Allan. 2001. Back in style: Reworking audience design. *Style and sociolinguistic variation*, ed. by Penelope Eckert and John R. Rickford, 139–69. Cambridge: Cambridge University Press.

Boersma, Paul, and David Weenink. 2002. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10).341–45.

Brysbaert, Marc, and Boris New. 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. *Behavior Research Methods* 41(4).977–90. DOI: 10.3758/BRM.41.4.977.

Campbell-Kibler, Kathryn. 2012. The implicit association test and sociolinguistic meaning. *Lingua* (Special issue: *New horizons in sociophonetic variation and change*, ed. by Nanna Haug Hilton, Charlotte Gooskens, Alexandra N. Lenz, and Anja Schüppert) 122(7).753–63. DOI: 10.1016/j.lingua.2012.01.002.

Campbell-Kibler, Kathryn. 2016. Towards a cognitively realistic model of meaningful sociolinguistic variation. *Awareness and control in sociolinguistic research*, ed. by Anna M. Babel, 123–51. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139680448.008.

Crowne, Douglas P., and David Marlowe. 1960. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 24(4).349–54. DOI: 10.1037/h0047358.

Dinkin, Aaron J., and Robin Dodsworth. 2017. Gradience, allophony, and chain shifts. *Language Variation and Change* 29(1).101–27. DOI: 10.1017/S0954394517000035.

Dodsworth, Robin. 2014. Network embeddedness and the retreat from Southern vowels in Raleigh. *University of Pennsylvania Working Papers in Linguistics* 20(2).41–50. Online: https://repository.upenn.edu/pwpl/vol20/iss2/6.

Dodsworth, Robin, and Mary Kohn. 2012. Urban rejection of the vernacular: The SVS undone. *Language Variation and Change* 24(2).221–45. DOI: 10.1017/S0954394512000105.

D'Onofrio, Annette. 2015. Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics* 19(2).241–56. DOI: 10.1111/josl.12115.

Drager, Katie; Jennifer Hay; and Abby Walker. 2010. Pronounced rivalries: Attitudes and speech production. *Te Reo* 53.27–53.

Fasold, Ralph W. 1972. *Tense marking in Black English: A linguistic and social analysis*. (Urban language series 8.) Arlington, VA: Center for Applied Linguistics. Online: https://eric.ed.gov/?id=ED129065.

Foulkes, Paul, and Gerard Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34(4).409–38. DOI: 10.1016/j.wocn.2005.08.002.

Fowler, Carol A. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14(1).3–28. DOI: 10.1016/S0095-4470(19)30607-2.

Galantucci, Bruno; Carol A. Fowler; and M. T. Turvey. 2006. The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review* 13(3).361–77. DOI: 10.3758/BF03193857.

Gambi, Chiara, and Martin J. Pickering. 2013. Prediction and imitation in speech. *Frontiers in Psychology* 4:340. DOI: 10.3389/fpsyg.2013.00340.

GARRETT, ANDREW, and KEITH JOHNSON. 2013. Phonetic bias in sound change. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 51–97. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199573745.003.0003.

GENTILUCCI, MAURIZIO, and PAOLO BERNARDIS. 2007. Imitation during phoneme production. *Neuropsychologia* 45(3).608–15. DOI: 10.1016/j.neuropsychologia.2006.04.004.

GILES, HOWARD; NIKOLAS COUPLAND; and JUSTINE COUPLAND. 1991. Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, ed. by Howard Giles, Justine Coupland, and Nikolas Coupland, 1–68. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511663673.001.

GOLDINGER, STEPHEN D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2).251–79. DOI: 10.1037/0033-295x.105.2.251.

GOLDINGER, STEPHEN D., and TAMIKO AZUMA. 2004. Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review* 11(4).716–22. DOI: 10.3758/BF03196625.

GUY, GREGORY R. 2013. The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics* 52.63–71. DOI: 10.1016/j.pragma.2012.12.019.

GUY, GREGORY R., and FRANS HINSKENS. 2016. Linguistic coherence: Systems, repertoires and speech communities. *Lingua* 172–173.1–9. DOI: 10.1016/j.lingua.2016.01.001.

HALL, JOSEPH SARGENT. 1942. The phonetics of Great Smoky Mountain speech. *American Speech* 17(2.2).1–110. DOI: 10.2307/487132.

HAY, JENNIFER, and KATIE DRAGER. 2010. Stuffed toys and speech perception. *Linguistics* 48(4).865–92. DOI: 10.1515/ling.2010.027.

JOHN, OLIVER P.; EILEEN M. DONAHUE; and ROBERT L. KENTLE. 1991. *The Big Five Inventory—versions 4a and 54*. Berkeley: Institute of Personality and Social Research, University of California, Berkeley.

JOHNSON, KEITH. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34(4).485–99. DOI: 10.1016/j.wocn.2005.08.004.

KIM, DONGHYUN, and MEGHAN CLAYARDS. 2019. Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience* 34(6).769–86. DOI: 10.1080/23273798.2019.1582787.

KIM, JONNY, and KATIE DRAGER. 2017. Sociophonetic realizations guide subsequent lexical access. *Proceedings of Interspeech 2017*, 621–25. DOI: 10.21437/Interspeech.2017-1742.

KUZNETSOVA, ALEXANDRA; PER B. BROCKHOFF; and RUNE H. B. CHRISTENSEN. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13).1–26. DOI: 10.18637/jss.v082.i13.

LABOV, WILLIAM. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

LABOV, WILLIAM. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.

LABOV, WILLIAM. 2010. *Principles of linguistic change, vol. 3: Cognitive and cultural factors*. Oxford: Wiley-Blackwell.

LABOV, WILLIAM; SHARON ASH; and CHARLES BOBERG. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

LEWANDOWSKI, NATALIE, and MATTHIAS JILKA. 2019. Phonetic convergence, language talent, personality, and attention. *Frontiers in Communication* 4:18. DOI: 10.3389/fcomm.2019.00018.

LIBERMAN, ALVIN M., and IGNATIUS G. MATTINGLY. 1985. The motor theory of speech perception revised. *Cognition* 21(1).1–36. DOI: 10.1016/0010-0277(85)90021-6.

LOBANOV, BORIS. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49.606–8. DOI: 10.1121/1.1912396.

LOUDERMILK, BRANDON C. 2015. Implicit attitudes and the perception of sociolinguistic variation. *Responses to language varieties: Variability, processes and outcomes*, ed. by Alexei Prikhodkine and Dennis R. Preston, 137–56. Amsterdam: John Benjamins.

LOVE, JESSICA, and ABBY WALKER. 2013. Football versus football: Effect of topic on /r/ realization in American and English sports fans. *Language and Speech* 56(4).443–60. DOI: 10.1177/0023830912453132.

Miller, Rachel M.; Kauyumari Sanchez; and Lawrence D. Rosenblum. 2010. Alignment to visual speech information. *Attention, Perception and Psychophysics* 72(6).1614–25. DOI: 10.3758/APP.72.6.1614.

Mitterer, Holger, and Mirjam Ernestus. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109(1).168–73. DOI: 10.1016/j.cognition.2008.08.002.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2).132–42. DOI: 10.1016/j.wocn.2010.12.007.

Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119(4).2382–93. DOI: 10.1121/1.2178720.

Pardo, Jennifer S.; Adelya Urmanche; Sherilyn Wilman; Jaclyn Wiener; Nicholas Mason; Keagan Francis; and Melanie Ward. 2018. A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics* 69.1–11. DOI: 10.1016/j.wocn.2018.04.001.

Pickering, Martin J., and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(2).169–90. DOI: 10.1017/S0140525X04000056.

Pickering, Martin J., and Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36(4).329–92. DOI: 10.1017/S0140525X12001495.

Podlipský, Václav Jonáš, and Šárka Šimáčková. 2015. Phonetic imitation is not conditioned by preservation of phonological contrast but by perceptual salience. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0399.pdf.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: https://www.R-project.org.

Rácz, Péter; Jennifer B. Hay; and Janet B. Pierrehumbert. 2017. Social salience discriminates learnability of contextual cues in an artificial language. *Frontiers in Psychology* 8:51. DOI: 10.3389/fpsyg.2017.00051.

Rácz, Péter; Jennifer B. Hay; and Janet B. Pierrehumbert. 2020. Not all indexical cues are equal: Differential sensitivity to dimensions of indexical meaning in an artificial language. *Language Learning* 70(3).848–85. DOI: 10.1111/lang.12402.

Reed, Paul E. 2014. Inter- and intra-generational monophthongization and southern Appalachian identity. *Southern Journal of Linguistics* 38(1).159–94.

Reed, Paul E. 2016. *Sounding Appalachian: /aɪ/ monophthongization, rising pitch accents, and rootedness*. Columbia: University of South Carolina dissertation. Online: https://scholarcommons.sc.edu/etd/3563/.

Rickford, John R., and Faye McNair-Knox. 1994. Addressee- and topic-influenced style shift: A quantitative sociolinguistic study. *Sociolinguistic perspectives on register*, ed. by Douglas Biber and Edward Finegan, 235–76. Oxford: Oxford University Press.

Sanchez, Kauyumari; Jennifer Hay; and Elissa Nilson. 2015. Contextual activation of Australia can affect New Zealanders' vowel productions. *Journal of Phonetics* 48.76–95. DOI: 10.1016/j.wocn.2014.10.004.

Tamminga, Meredith. 2019. Interspeaker covariation in Philadelphia vowel changes. *Language Variation and Change* 31(2).119–33. DOI: 10.1017/S0954394519000139.

Thakerar, Jitendra N.; Howard Giles; and Jenny Cheshire. 1982. Psychological and linguistic parameters of speech accommodation theory. *Advances in the social psychology of language*, ed. by Colin Fraser and Klaus R. Scherer, 205–55. Cambridge: Cambridge University Press.

Theodore, Rachel M., and Joanne L. Miller. 2010. Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America* 128(4).2090–99. DOI: 10.1121/1.3467771.

Theodore, Rachel M.; Joanne L. Miller; and David DeSteno. 2009. Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America* 125(6).3974–82. DOI: 10.1121/1.3106131.

Thomas, Erik R. 2011. Sociolinguistic variables and cognition. *WIREs Cognitive Science* 2(6).701–16. DOI: 10.1002/wcs.152.

TORBERT, BENJAMIN. 2010. The salience of two Southern vowel variants: Fronted /o/ and weak-glided /aɪ/. *Southern Journal of Linguistics* 34(2).1–36.
TRUDGILL, PETER. 1986. *Dialects in contact*. Oxford: Basil Blackwell.
VAN BERKUM, JOS J. A.; DANIELLE VAN DEN BRINK; CATHELIJNE M. J. Y. TESINK; MIRIAM KOS; and PETER HAGOORT. 2008. The neural integration of speaker and message. *Journal of Cognitive Neuroscience* 20(4).580–91. DOI: 10.1162/jocn.2008.20054.
VAUGHN, CHARLOTTE, and TYLER KENDALL. 2019. Stylistically coherent variants: Cognitive representation of social meaning. *Revista de Estudos da Linguagem* 27(4).1787–1830. DOI: 10.17851/2237-2083.0.0.1787-1830.
WADE, LACEY. 2020. *The linguistic and the social intertwined: Linguistic convergence toward Southern speech*. Philadelphia: University of Pennsylvania dissertation. Online: https://repository.upenn.edu/dissertations/AAI27957422.
WADE, LACEY, and GARETH ROBERTS. 2020. Linguistic convergence to observed versus expected behavior in an alien-language map task. *Cognitive Science* 44(4):e12829. DOI: 10.1111/cogs.12829.
WALKER, ABBY. 2019. The role of dialect experience in topic-based shifts in speech production. *Language Variation and Change* 31(2).135–63. DOI: 10.1017/S0954394519000152.
WALKER, ABBY, and KATHRYN CAMPBELL-KIBLER. 2015. Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology* 6: 546. DOI: 10.3389/fpsyg.2015.00546.
WEISS, NEIL A. 2015. wPerm: Permutation tests. R package version 1.0.1. Online: https://cran.r-project.org/package=wPerm.
WILSON, COLIN; ELEANOR CHODROFF; and KUNIKO NIELSEN. 2016. Generalization in VOT imitation: Feature adaptation or acoustic-phonetic covariation? *The Journal of the Acoustical Society of America* 140.3344. DOI: 10.1121/1.4970684.
WOLFRAM, WALT, and DONNA CHRISTIAN. 1976. *Appalachian speech*. Arlington, VA: Center for Applied Linguistics.
YU, ALAN C. L. 2010. Perceptual compensation is correlated with individuals' 'autistic' traits: Implications for models of sound change. *PLoS ONE* 5(8):e11950. DOI: 10.1371/journal.pone.0011950.
YU, ALAN C. L.; CARISSA ABREGO-COLLIER; and MORGAN SONDEREGGER. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and 'autistic' traits. *PLoS ONE* 8(9):e74746. DOI: 10.1371/journal.pone.0074746.
YUAN, JIAHONG, and MARK LIBERMAN. 2008. Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics 2008*, 5687–90.
ZEHR, JÉRÉMY, and FLORIAN SCHWARZ. 2018. PennController for Internet Based Experiments (IBEX). DOI: 10.17605/OSF.IO/MD832.
ZELLOU, GEORGIA; DELPHINE DAHAN; and DAVID EMBICK. 2017. Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition and Neuroscience* 32(6).776–94. DOI: 10.1080/23273798.2016.1275710.
ZELLOU, GEORGIA; REBECCA SCARBOROUGH; and KUNIKO NIELSEN. 2013. Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America* 140(5).3560–75. DOI: 10.1121/1.4966232.

[laceya@sas.upenn.edu]

[Received 26 August 2020;
revision invited 25 October 2020;
revision received 5 April 2021;
accepted pending revisions 10 July 2021;
revision received 5 September 2021;
accepted 13 September 2021]