GLOSSA psycholinguistics

Wade, L., Embick, D., & Tamminga, M. (2023). Dialect experience modulates cue reliance in sociolinguistic convergence. *Glossa Psycholinguistics*, 2(1): 19, pp. 1–30. DOI: https://doi.org/10.5070/G6011187

eScholarship University of California

Dialect experience modulates cue reliance in sociolinguistic convergence

Lacey Wade, University of Kansas, US, laceywade@ku.edu
David Embick, University of Pennsylvania, US, embick@ling.upenn.edu
Meredith Tamminga, University of Pennsylvania, US, tamminga@ling.upenn.edu

Expectation-driven convergence occurs when speakers shift their speech to approximate the language they expect, rather than observe, from their interlocutor. In Wade (2022), participants produced more monophthongal /aɪ/ – a salient feature of Southern U.S. English – after hearing other Southern-accented features. Here, by decoupling acoustic and social information with a dialect-label manipulation task, we investigate what types of cognitive associations account for this behavior: indirect socially-mediated associations that rely on recognizing that monophthongal /aɪ/ and other Southern-accented variants are both associated with the social category "Southern," or direct associations between variants that rely on tracking their common co-occurrence at the individual level. We find that both acoustic and social-label cues trigger convergence, but in-group speakers from the South rely on acoustic cues, while out-group speakers from outside of the South are best cued by social-category labels. Results indicate a crucial role of dialect experience in the encoding and utilization of sociolinguistic knowledge.

1. Introduction

Listeners track information about both group and individual patterns of language usage. Providing listeners with information about a social category that a talker belongs to can shape the way language is processed. For instance, beliefs about a talker's age (Koops et al., 2008), gender (Strand, 1999), race (Staum Casasanto, 2010), regional origins (D'Onofrio, 2015; Hay et al., 2006; Niedzielski, 1999), and other socio-demographic characteristics have been shown to alter the way the same linguistic stimulus is categorized. At the same time, people track information about *individual speakers'* linguistic patterns. Work on perceptual learning has shown that listeners adjust their perceptual category boundaries between two phonemes when exposed to a novel pronunciation. For example, training listeners that a talker produces /s/ ambiguously between [s] and [ʃ] (so that *sip* sounds ambiguous between *sip* and *ship*) leads to a shift in the listeners' s/ʃ boundary when categorizing tokens from the same talker. When categorizing tokens from a different talker, however, the same perceptual boundary shift is often weaker (Lai, 2021) or absent (Eisner & McQueen, 2005; Kraljic & Samuel, 2005), suggesting that people retain and utilize information about talker-specific variability. Likewise, Zellou, Dahan, and Embick (2017) showed that speakers track talker-specific information in a shadowing task eliciting phonetic convergence. Participants converged toward the average degree of coarticulatory nasalization a talker produced across blocks of greater or lesser nasalization, suggesting they tracked talker-specific nasalization throughout the course of the experiment.

Patterns observed at both the group and individual levels can also influence processing in a *new* situation, such as with a novel talker. Perhaps most obviously, in order to generate a relevant group label (such as "people from Pittsburgh") and associate it with a linguistic variant commonly observed among that group (such as "yinz" as a second person plural pronoun), a listener must abstract from various encounters with Pittsburghers to form a mental category for "people from Pittsburgh."[1] People can use this information when processing speech from novel talkers about whom they have no linguistic information (yet), as long as they know something about the social category to which the talker belongs. In a similar vein, generalization may occur with information tracked at the individual level *without reference to a particular social category*. Listeners may observe from multiple speakers that people with longer voice onset time (VOT) for word-initial /p/ also have longer VOT for /t/ and /k/, even without there being a shared social category for long-VOT speakers. After enough accumulated individual-speaker encounters indicating covariation of VOT among voiceless stops, this information may generalize to new talkers. Hearing one component of an individual's linguistic system (e.g., long VOT /p/) can

[1] Though note that this information could also plausibly be passed explicitly through interpersonal communication rather than implicitly through direct observation, in which case it may come in an already somewhat generalized form (e.g., "Pittsburghers say *yinz*").

lead to predictions about some other component that has not been observed (e.g., long VOT /k/) (Theodore & Miller, 2010). While there is evidence that individual VOT covariation is a consequence of structural constraints on, e.g., a shared [+spread glottis] feature (Chodroff & Wilson, 2017), we expect tracking of covariation at the individual level could occur in the absence of such structural relationships as well.

In some cases, particularly when observation of one linguistic variant initiates a response targeting some other variant, it may not be clear what type of information a listener (or speaker, as we will see) is utilizing. One example is when listeners expect to observe *socio-stylistic coherence* among two or more linguistic variables. For instance, Campbell-Kibler (2012) found that the *-in* variant of variable (ING) was associated with monophthongal /aɪ/ in an Implicit Association Test, suggesting this is because both variants are consistent with a "Southern" speech style. Vaughn and Kendall (2019) found that the same association between the *-in* variant and monophthongal /aɪ/ impacted *production* as well. Participants who were instructed to produce sentences with the *-in* variant also shifted their production of /aɪ/ to be more monophthongal, which they suggest is due to the stylistic coherence of these two variants. Wade's (2022) observation that speakers exhibit *expectation-driven convergence* is another example. Participants who heard a Southern-accented talker give clues in a Word Naming Game produced more monophthongal /aɪ/ vowels in their spoken responses, even though the Southern-accented talker never produced this vowel. No such shift was observed when participants heard a Midland-accented talker from Ohio. Wade concluded that participants converged toward a linguistic form they reasonably expected, but did not directly observe, from the Southern-accented talker, referring to this as *expectation-driven convergence*. Both expectation-driven convergence and style coherence provide evidence that expectations shape linguistic processing and even production – and that expectations for one linguistic variant can be cued by observations of *another related variant*. However, the cognitive mechanisms by which this behavior occurs are not well understood.

There are at least two different associative "routes" through which behaviors relying on associations between variants might arise, stemming from reliance on individual-level vs. group-level linguistic pattern tracking that we laid out above. One way of categorizing these different routes is whether linguistic expectations are triggered in a *top-down* manner, beginning with information about a talker's social group membership, vs. a *bottom-up* (or perhaps bottom-across?) manner, beginning with lower-level acoustic cues. Here we refer to these as (1) an **indirect socially-mediated** route relying on knowledge that variants share a social meaning or are used in stylistically congruent ways or by the same social groups, and (2) a **direct co-occurrence** route relying on statistical knowledge that variants tend to co-occur (i.e., within individuals or perhaps even temporal chunks). Under a direct co-occurrence route, variants that commonly co-occur (i.e., at the speaker level) form associative links in the mind that allow for one variant to activate the other directly. Under an indirect, socially mediated route, the same sort of associative link

exists, but rather than directly linking variants to other variants, it is mediated by a shared social category. Vaughn and Kendall (2019) describe indirect social mediation with an analogy about clothing style:

> Consider a person, say a young American man in particular, getting dressed and deciding to wear a salmon pink polo shirt (analogous here to producing –in). That shirt decision may invoke the style "preppy", which calls to mind other articles of clothing also associated with a "preppy" style. For instance, the dresser may then decide to wear khaki pants and loafers (wearing articles of clothing that co-occur stylistically)...it is possible to initiate the style "preppy", and subsequent fashion choices, by an initial shirt selection. Producing –in/wearing a salmon pink shirt may call to mind a Southern/preppy style, which can affect other variant/clothing selections. (Vaughn & Kendall, 2019, p.1809)

Using the same analogy, under a direct co-occurrence route, this man might decide to complement his salmon pink polo shirt with a pair of khaki pants not because they invoke a "preppy" style necessarily, but because he is accustomed to wearing, or seeing other people wear, these items together. Perhaps he even purchased this shirt and pants combination after seeing it displayed on a mannequin in a shop window. In other words, he wears these items together because he has seen them together before.

Both socially-mediated and co-occurrence routes are independently attested in work on speech processing and production.[2] For one, providing participants with social information about a talker has been shown to influence speech processing. Niedzielski (1999) found that telling listeners a speaker was from Canada or Detroit influenced their responses on a vowel-matching task. When asked to choose the MOUTH-class token they just heard, participants were more likely to choose a raised token (consistent with Canadian raising) when they were told the talker was from Canada. Hay, Nolan, and Drager (2006) conducted a similar study, where participants listened to a New Zealand English speaker reading a list of sentences, then selected a token from a synthesized continuum ranging from *fush* to *feesh* that best matched the /ɪ/ vowels they heard. When "Australian" was printed on top of participants' answer sheets, they selected higher and fronter tokens than when "New Zealand" was printed on the answer sheet, consistent with production differences across these dialects. Further, D'Onofrio (2018) found that participants who were visually cued to expect a "Valley Girl" persona were more likely to

[2] Though note that the socially-mediated route is attested when the triggering event is a social category/label but not when it is linguistic. We don't know how much of phenomena like stylistic coherence and expectation-driven convergence are socially-mediated, but we know social triggers exist. Because the relationship between social knowledge and linguistic knowledge is bi-directional (i.e., social information can trigger linguistic expectations, and linguistic information can trigger social expectations), when we combine these, we get a socially-mediated linkage between two variants.

consider a backed-TRAP variant to be a production of the TRAP class, as TRAP-backing is used in constructing a "Valley Girl" persona. These are just a few of many such studies showing that social expectations influence the way the same speech sound is processed and categorized.

We also have evidence that co-occurring linguistic cues influence expectations for each other. For instance, Theodore and Miller (2010) found that participants developed predictions about a talker's VOT for one voiceless stop based on knowledge of their VOT use for a different voiceless stop. Similarly, Nielsen (2011) found that participants who shadowed a talker with lengthened VOT for /p/-initial words went on to imitate lengthened VOT, not only for /p/-initial words, but also for /k/-initial words, which were not heard during the experiment. (Kraljic & Samuel, 2006) likewise found that perceptual learning of the /d/-/t/ phonemic boundary generalized to /b/-/p/.[3]

Here, we ask what happens when co-occurrence and social category-mediated associations may both be plausibly utilized, focusing specifically on cases of *expectation-driven convergence* where hearing certain variants cues convergence toward a socio-stylistically related variant. We take as our starting point Wade's (2022) study showing that people exhibit *expectation-driven convergence* toward monophthongal /aɪ/, a feature of Southern U.S. English. The Southern dialect is one of the most recognizable regional dialects in North America, consisting of features such as reversed tense and lax front vowel nuclei, fronted back vowels, and PIN-PEN merger. Monophthongal /aɪ/ is a particularly socially salient feature associated with Southern speech (e.g., Hall, 1942; Labov, 2010; Reed, 2014; Wolfram & Christian, 1976). When /aɪ/ is monophthongized in the South, it is produced closer to [a:], with the primary change being in the weakening or omission of the glide (Labov, 1994). While /aɪ/ monophthongization occurs in all phonological environments in some areas of the South, such as the inland South, in other areas of the South it occurs only before voiced segments or in coda position. /aɪ/ monophthongization, and the Southern Vowel Shift more generally, is retreating among younger speakers, particularly in urban areas in the South (Dodsworth & Kohn, 2012). Therefore, many Southerners do not exhibit stereotypical Southern-accented features.

Wade (2022) initially examined this variant due to its social salience and found that participants who heard a Southern accent (without words containing /aɪ/) spontaneously converged toward that accent by producing more monophthongal /aɪ/ vowels, even though they never actually observed how the talker produced the /aɪ/ vowel. These results are reproduced in **Figure 1**. This link between the Southern-accented variants participants observed and then went on to produce (monophthongal /aɪ/) may have reasonably been accomplished through either a direct co-occurrence route or an indirect socially-mediated route. Here, we replicate

[3] It is worth noting that these studies all focus on VOT, so we know little about the possibility of direct co-occurrence among other co-varying features.

Wade's experiment but isolate social category cues and acoustic cues to determine which route(s) participants rely on in this type of related-variant-activation behavior.

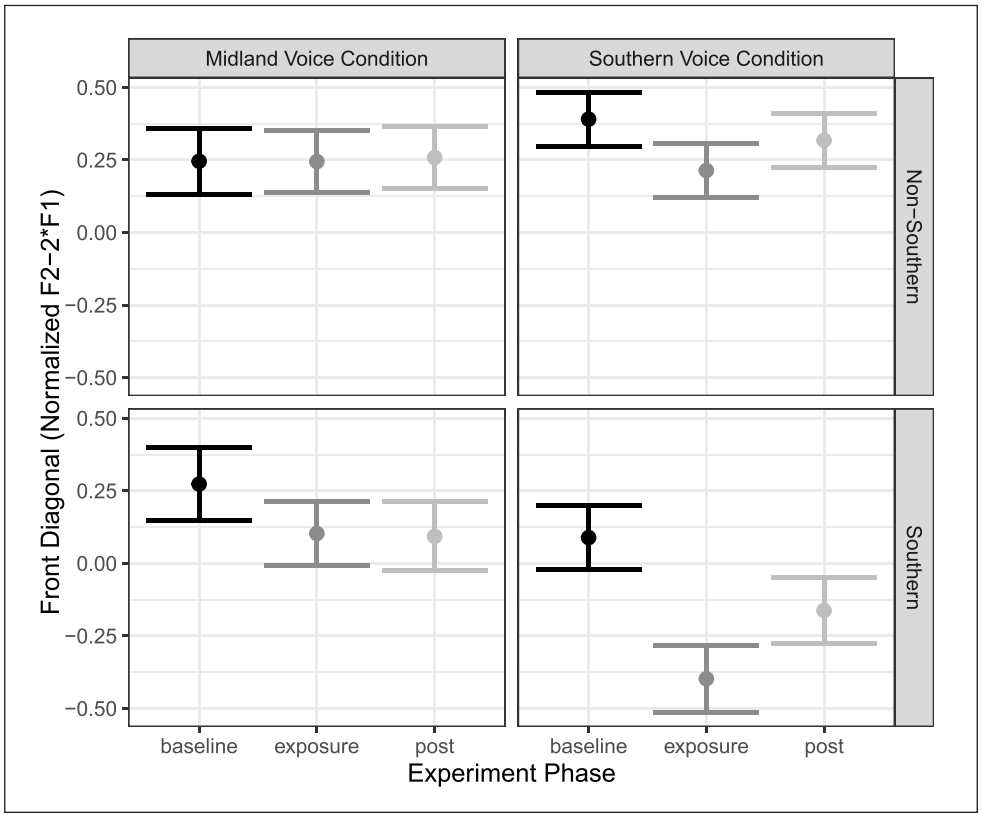

Figure 1: Shift across phases by Voice Condition and participant Dialect. Mean values with 95% confidence intervals for all tokens. Reproduced from Wade (2022), Experiment 2.

We introduce a dialect-label mismatch manipulation into the Word Naming Game paradigm developed in Wade (2020) and Wade (2022). The Word Naming Game prompts participants to produce particular words with "clues," and convergence is determined by comparing production of target words in a baseline (pre-exposure) phase to those produced during exposure to a Southern-accented model talker who – crucially – never produces the target /aɪ/ vowel. Since our goal here is to tease apart potential triggers for expectation-driven convergence, we assign incongruent dialect labels to the model talkers, such that, in each condition, only one of these sources of information would cue shifts toward Southern speech. Previous work has confirmed that social expectations may override linguistic information (Carmichael, 2018; Niedzielski, 1999; Williams, 1989); here, we ask under which conditions participants may rely on one type of information over another.

We compare two conditions, where voice and social labels are incongruent:[4] In the "Voice-Triggered" Condition, participants hear a Southern-accented talker but are told the talker is from the Midwest and has a "Midwest accent,"[5] so only acoustic information (e.g., other Southern-accented vowels besides /aɪ/) would provide cues to activate monophthongal /aɪ/. In the "Label-Triggered" Condition, participants hear a Midland-accented talker but are told the talker is from the South and has a Southern accent; in this case, only explicit social information would provide cues to activate monophthongal /aɪ/. If the link between observed variants and monophthongal /aɪ/ is primarily mediated by a shared social label, the actual voice of the talker should have little effect, since the social category of the model talker is already known. However, if observed variants are directly linked to monophthongal /aɪ/ due to their common co-occurrence, we expect to observe an independent effect of talker voice, such that a Southern voice would elicit convergence, even in the absence of explicit social labels indicating Southernness – or indeed, even in the presence of a contradictory social label. Crucially, participants never hear any tokens of /aɪ/ from the model talker in either the exposure phase or auditory instructions of either condition.

We make the following predictions:

1. If participants primarily rely on an indirect socially-mediated associative route, we expect to observe convergence toward the Southern label.
2. Alternatively, if participants primarily rely on a direct co-occurrence associative route, we expect to observe convergence toward the Southern voice.
3. If participants equally rely on a combined direct and indirect route, we predict this may manifest as either (a) convergence toward both the Southern label and the Southern voice or (b) no convergence, indicating contradictory pieces of information overriding each other. Regardless, we expect to see no difference in convergence patterns across conditions.

We crucially ask whether these behaviors differ based on the dialect background of the participant. In Wade's (2022) study, Southerners were shown to converge significantly more than Non-Southerners. If the reason for this was that Southerners were better able to recognize a talker as Southern based on their voice alone (no social labels were provided in this study), we would expect that Southerners and Non-Southerners may behave more similarly to one another when

[4] We avoid directly comparing congruent and incongruent conditions within this experiment because, as, e.g., McGowan (2015) has shown, (in)congruency of voices and social attributes impacts speech processing. We therefore view our two incongruent conditions as more directly comparable.

[5] The term "Midwest" was used in the experiment materials because we expected participants would be more likely to recognize the term "Midwest" (compared to Midland) as referring to the Midland dialect.

given explicit social label information in the present study. However, an alternative possibility is that Southerners in Wade (2022) converged more because they generally rely more on talker voice cues, while Non-Southerners might rely more on social labels, which were absent in that study. We have some reason to believe that Non-Southerners may in fact rely more heavily on social labels than Southerners. Previous work has suggested that out-group members are more likely to rely on stereotyped associations in sociolinguistic perception, while in-group members rely more on their more nuanced and veridical past experiences. For instance, Drager and Kirtley (2016) show that people without military experience interpreted Southern speech as sounding "military," due to often stereotyped portrayals of military speech in television and film – their main source of exposure to military speech – while those with military experience did not make such associations. We therefore might expect Non-Southerners, who have less experience with Southern speech, to similarly rely on top-down dialect labels linking Southernness to a highly salient and stereotyped feature: monophthongal /aɪ/. Southerners, on the other hand, may rely on a more accurate indicator of monophthongal /aɪ/: presence of other Southern-accent features. We test this prediction by directly comparing which cues trigger convergence for Southerners compared to Non-Southerners.

2 Materials and methods

2.1 Participants

Participants (N = 120) were recruited from Prolific to participate in this experiment and were given ~$10 for their time. Of these, 2 were excluded for giving incorrect responses to more than 50/180 clues, leaving 118 participants analyzed here. Of these, 114 disclosed demographic information. Seventy-three identified as female, 39 as male, and 2 as non-binary. In regard to maximum educational attainment, 63 reported some college, 14 a Bachelor's degree, 19 a high school diploma, 11 an associate's degree, 6 an advanced or graduate degree, and 2 other. 68 identified as White, 14 as Black or African American, 12 as Asian or Pacific Islander, 10 as Hispanic or Latinx, 3 as Native American or American Indian, and 7 as a racial background not listed. Mean age was 23.4, with a range of 18–58. Participants were all American English speakers and reported no speech or hearing impairments. Fifty-three participants were from the U.S. South, while the remaining 65 were from the U.S. but outside of the South, as shown in **Table 1**. If participants spent the majority of their school-aged years, ages 5–18, within the Southern isogloss identified in Labov, Ash, and Boberg (2006), they were labeled as "Southern." Otherwise, they were labeled 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. All participants spent the vast majority of their school-aged years in the United States.

Table 1: Participant breakdown by experimental condition (Southern Voice vs. Southern Label) and participant dialect background (Southern vs. Non-Southern).

	Southern Voice	**Southern Label**
Southerners	26	27
Non-Southerners	32	33

2.2 Procedure

The study was conducted under the University of Pennsylvania's Institutional Review Board, Protocol #820633. The experiment was implemented using PCIbex (Zehr & Schwarz, 2018) and administered via participants' web browsers, and participants were recorded through their computer microphones. After giving informed consent, participants tested their microphone and headphones, then began the experiment with a demographic survey collecting age, gender identity, race/ethnicity, educational attainment, and residential history information, the latter of which was used for sorting participants into two broad dialect-background categories: Southerners and Non-Southerners.

Next, participants completed the Word Naming Game task, where clues were given to participants who were instructed to state the word described by each clue, using the carrier phrase "The word is __." For example, a clue might read "The saying goes, if at first you don't succeed, do this three letter T-word again," and participants would respond into their microphone, "The word is *try*." The first 60 clues were presented orthographically on-screen to collect participants' baseline responses prior to any exposure. In the exposure phase, the next 60 clues were presented auditorily, as read by one of our model talkers, either a Southern-accented talker or a Midland-accented talker, depending on the condition to which the participant was randomly assigned. Acoustic measurements of participant's vowel productions in the Exposure Phase, compared to those in the Baseline Phase, are used as a measure of convergence. When participants were ready to record their response, they pressed a "Record" button on their screen, and the screen indicated that they were recording. After recording their response, participants pressed the "Next" button to automatically save their recorded response and continue to the next clue, and the screen indicated that they were no longer recording. Since a pilot version of the experiment showed that incorrect responses led to significant data loss, the Word Naming Game task also included a "hint" for each clue, which provided the number of letters in the correct response, indicated by blank spaces, with 1–3 letters filled in (e.g., *t _ _ r*). An example of the experiment screen during a trial is shown in **Figure 2**. A perception task, followed by a post-exposure production task, were also administered to participants after the Exposure Phase, but these components are beyond the scope of this paper and will not be discussed here.

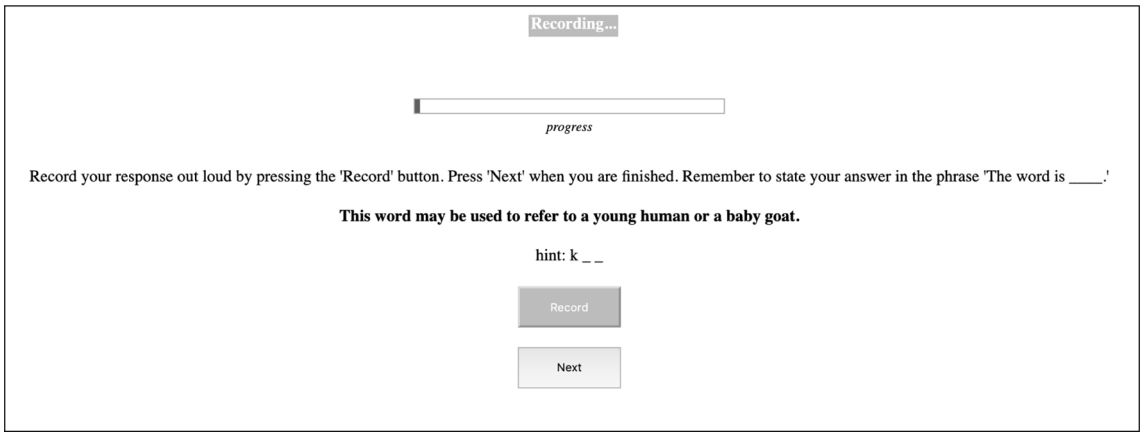

Figure 2: Screenshot of a trial in the baseline phase, where clues are presented on-screen.

At the beginning of the experiment, directions were presented 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.'*

At the beginning of the exposure phase, participants were introduced to the model talker they would hear giving clues, who provided short instructions. In addition to the instructions, the model talkers also introduced themselves as being from a particular dialect region and specifically called attention to their accent. Rather than assigning the talkers dialect labels in the instructions of the experiment, the model talkers introduced their own dialect labels in their real accents so that participants were less likely to think that the labels were inaccurate and swapped by mistake. The model talkers also gave their "name" and a "fact" about themselves, ostensibly to assess whether "feeling connected" to the model talker by learning something about them helps participants in the task, though in actuality these facts were added to draw less attention to the model talkers' accent as a crucial manipulation in the study. As such, the Southern-Labeled talker and the Midland-Labeled talker had slightly different scripts to make their calling attention to their accents seem less conspicuous by way of appearing more relevant to the task (e.g., suggesting that a Southern accent may lead to difficulty interpreting the clues or suggesting that a Midland accent may be helpful for the task.) The scripts the model talkers read are as follows:

Voice-triggered Condition (Southern Voice, Midland Label) Script: *For the next section, we are going to change the game a bit. The clues will be given to you out loud. Please make sure your headphones are on and volume is at a comfortable level. With me being from Michigan, hopefully this Midwest accent will be clear and help you guess the words easily. Our goal with this experiment is to determine whether people are better at guessing words when they are reading alone or interacting with another person. We think interacting with another person will make you better at guessing words, especially if you feel connected to your partner in the game. So here's a little bit of information about me: Name's Jenn, From Ann Arbor, Michigan, Hobbies include baking and playing guitar. Could you press the record button then briefly state some information about yourself too? Great, thanks! Let's get started with the next section.*

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

At the end of the experiment, participants were asked to rate how likely it was that the talker they heard was from the Midwest/South on a scale of 0 (unlikely) to 100 (likely).

In each phase, 30 target words containing the /aɪ/ vowel in word-final or pre-voiced consonant contexts[6] and 30 non-/aɪ/ filler words were elicited. The set of 60 words elicited in each phase was selected randomly from one of three predetermined sets, and no set (and therefore no item) was repeated within the experiment. The phase in which each set appeared was counterbalanced across participants with individual items within sets randomized for each participant. The target items in each set were balanced for place of articulation of adjacent segments, as well as for lexical frequency, determined using the SUBTLEXus corpus (Brysbaert & New, 2009) LOG10CD measure. Each set of target items had the same mean lexical frequency and standard deviation.

The clues used to elicit participant responses were one or two sentences long and were constructed to give participants sufficient evidence for the dialect of the model talker, including

[6] This is the phonological environment in which monophthongization reliably occurs in the U.S. South. Some varieties, such as those spoken in the inland South, also monophthongize before voiceless segments, but for consistency we include no pre-voiceless contexts here.

stressed tokens of /i, ɪ, e, ɛ, æ, u, o/, which undergo the Southern Vowel Shift. Crucially, the /aɪ/ vowel was never used in any of the clues or in the auditory directions, so any shifts toward monophthongal /aɪ/ would be the result of expectations alone. In the baseline phase, participants silently read the clues presented on the screen to themselves before responding. In the exposure phase, audio clues were read by either a Midland or Southern talker, both white women in their 20s. The Southern talker was from Hurley, Mississippi and produced typical Southern dialect features, such as raised front lax vowel nuclei, fronting of back vowels, and PIN-PEN merger. The Midland talker was from northeast Ohio and lacked these features. Additional information about the talkers and stimuli can be found in Wade (2020, 2022).

2.3 Analysis

Recordings were forced-aligned using the Penn Phonetics Lab Forced Aligner (p2fa) (Yuan & Liberman, 2008). Forced alignment was then blindly hand-checked and corrected as necessary, such that RAs correcting TextGrids did not know which condition or phase any given sound file was collected in. Praat (Boersma & Weenink, 2019) settings for max formant and number of formants were adjusted as necessary for each participant, and sometimes for each vowel. A Praat script was then used to automatically measure the first three formants of target vowels at 7 time points throughout the course of the vowel: 10%, 20%, 30%, 50%, 70%, 80%, and 90%. Formant measurements were then normalized in R using the Lobanov method. Any incorrect responses given were omitted from the analysis, and all data visualization and statistical modeling was done in R (R-Core-Team, 2020). The measurement of interest is the realization of the glide (at 80% of the duration of the vowel) along the front diagonal (F2-2*F1) of the vowel space, which was then z-scored within participant. Minimal data trimming was done prior to analysis. Outliers more than three standard deviations from by-participant means for each vowel class were omitted (N = 77/21,050 or 0.4% of all tokens)). A linear mixed effects regression model was fit to the data predicting this front diagonal measure, 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. The bobyqa optimizer was used to facilitate model convergence.

One large model with all possible fixed and random effects of interest would not converge, so we built several iterative models (publicly available on OSF) and performed model comparison using the step() function in the lmerTest package in R, which performs backward elimination of random effects followed by fixed effects. This function uses likelihood ratio tests for random effects and F-tests, using Satterthwaite's approximation, for fixed effects. Three-way interactions (Phase*Freq*Dialect) and (Phase*Freq*Condition) and a four-way interaction (Phase*Freq*Condition*Dialect) investigating the role of frequency in shifts across different conditions/groups did not significantly improve the model and are not included in the final

model, but Phase*Freq does, and is included. We also tested vowel duration in interaction with Phase*Dialect*Condition, but none of these interactions were significant, so duration was only retained as a main effect. Southernness ratings were also tested in a four-way interaction (Phase*Condition*Dialect* Southernness), but the addition of Southernness ratings (as a main effect or in interactions) did not significantly improve model fit, and are excluded here as well. Note that the inclusion of Southernness ratings in the model does not noticeably alter the results or the conclusions we might draw about them. Finally, inclusion of Order in which the set of lists was elicited did not improve the model as a main effect or in any subset of interactions in Order*Phase*Condition*Dialect, suggesting that the lists that were randomly elicited in either the baseline or exposure phase were roughly equivalent, and one did not contribute to greater shifts than another. Random intercepts were included for participant and word. All fixed effects (including interaction terms) were tested as random slopes by participant and word, as appropriate given the study design. Only those that significantly improved model fit are included: (Phase + Duration|Participant) and (Condition + Dialect + Duration|Word). Only the following fixed predictors are included in the final model and significantly improved model fit:

Phase:	Categorical predictor referring to the experimental phase, treatment coded with levels BASELINE (reference) and EXPOSURE.
Dialect:	Categorical predictor referring to participant dialect, treatment coded with levels SOUTHERN (reference) and NON-SOUTHERN.
Duration:	Continuous predictor referring to vowel duration, z-scored, meant to control for effects of global changes like speech rate shifts on /aɪ/ glide realization.
Frequency:	Continuous predictor referring to lexical frequency, as measured by the Subtlex LOG10CD measure.
Phase*Cond*Dial:	Three-way interaction intended to test whether dialect groups differ in which conditions elicit convergence.
Phase*Freq:	Two-way interaction intended to test whether convergence is greater for infrequent items, as has previously been observed for input-driven convergence (Babel, 2010; Goldinger, 1998; Goldinger & Azuma, 2004; Nielsen, 2011).

The model was run once, then observations with absolute standardized residuals exceeding 3 standard deviations were excluded (38 tokens/0.6% of observations) before running the final model presented here. The final model includes a total token count of 6808, which amounts to an average of 58 tokens per participant. Data and scripts are available at https://osf.io/s8wuh.

3. Results

Glide realizations across the two experiment phases are shown in **Figure 3**, faceted by condition and participant dialect background. A table of post-hoc comparisons within and across these facets, using the *emmeans* package version 1.7.0 (Lenth, 2021) in R (R-Core-Team, 2020), is shown in **Table 3**, derived from the model in **Table 2**. Findings show that Southerners and Non-Southerners both shift toward monophthongal /aɪ/ during exposure, but they are triggered by different cues. Non-Southerners shift from their baselines to produce weaker glides (lower and further back along the front diagonal of the vowel space) in the exposure phase, but only when the talker is labeled as Southern (facet A, Est = –0.21, $p < 0.01$). When Non-Southerners hear a Southern Voice that is not labeled as such, they fail to shift, producing roughly equivalent /aɪ/ glides in the baseline and exposure phases, evidenced by overlapping confidence intervals (facet B, Est = –0.13, $p = 0.117$).

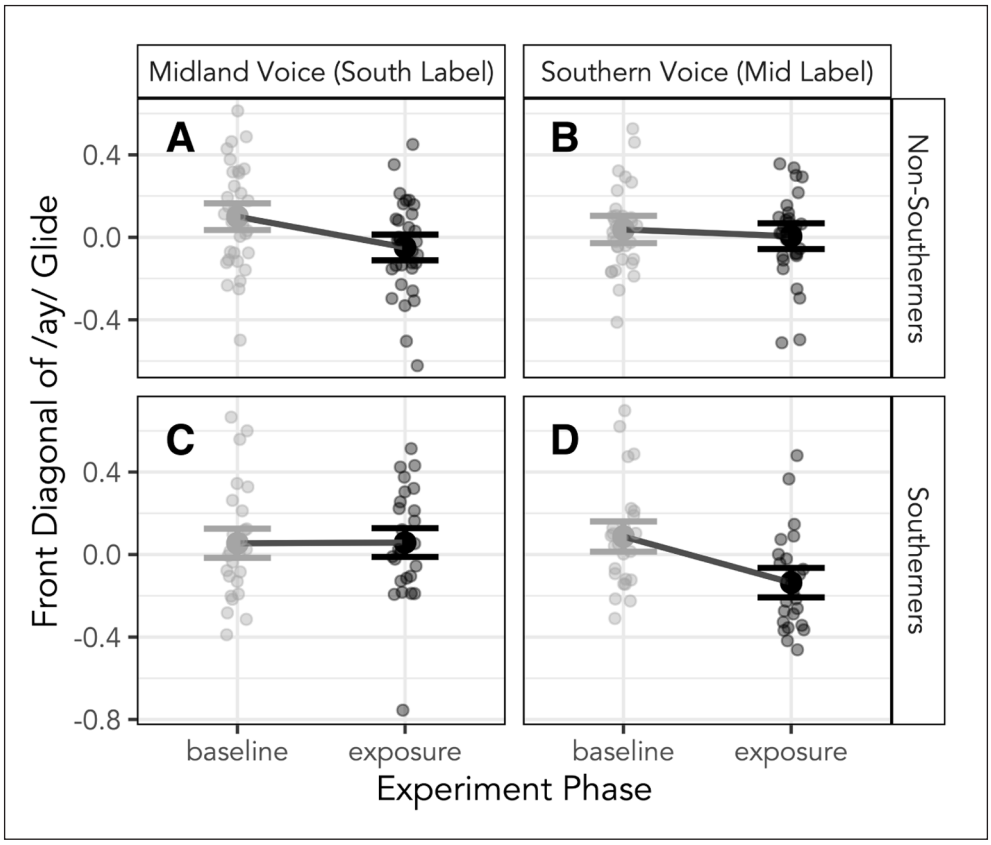

Figure 3: Shifts in /aɪ/ glide (80%) production from baseline to exposure, along the front diagonal of the vowel space (F2–2*F1), scaled within participant. Points represent speaker means. Error bars indicate 95% confidence intervals over all tokens.

Table 2: Table of fixed effects from linear model predicting front diagonal of the glide (80%). Random effects: (Phase + Duration|Participant) and (Condition + Dialect + Duration|Word).

	Estimate	SE	df	t-value	p-value	
(Intercept)	0.298	0.118	138.747	2.517	0.013	*
Phase (Exposure)	–0.262	0.089	188.306	–2.945	0.004	**
Condition (MidlandVoice)	–0.127	0.083	111.646	–1.527	0.130	
Dialect (NonSouthern)	–0.065	0.080	112.286	–0.812	0.418	
Duration	0.418	0.029	190.252	14.339	<0.0001	***
Frequency	–0.115	0.044	83.812	–2.590	0.011	*
Phase*Condition	0.266	0.111	115.251	2.390	0.018	*
Phase*Dialect	0.132	0.105	113.429	1.251	0.213	
Condition*Dialect	0.250	0.109	104.159	2.280	0.025	*
Phase*Frequency	0.058	0.021	5666.014	2.799	0.005	**
Phase*Condition*Dialect	–0.346	0.148	114.663	–2.328	0.022	*

Southerners, on the other hand, show the opposite pattern. When exposed to a Midland Voice that is labeled as a Southern talker, Southerners show no sign of shifting (facet C, Est = 0.003, p = 0.969). Baseline and exposure productions are not noticeably different, with overlapping confidence intervals. However, when Southerners are exposed to a Southern-accented talker whom they are told is from the Midwest, they shift from their baselines to produce weaker /aɪ/ glides in the exposure phase (facet D, Est = –0.262, p = 0.003). These shifts are roughly equivalent in size to the shifts Non-Southerners exhibit in the Southern Label condition. Comparisons across conditions are shown in **Figure 3**. Southerners' lack of convergence in the Southern Label condition (facet C) significantly differs from their shifts in the Southern Voice condition (facets C vs. D, Est = –0.266, p = 0.017), as well as Non-Southerners' shifts in the Southern Label condition (facets A vs. C, Est = 0.214, p = 0.041). However, the difference between Non-Southerners' lack of shift in the Southern Voice condition (facet B) and the other conditions that *do* show convergence (facets A and D) does not reach statistical significance.

These results may be compared to the original findings of Wade (2020), shown in **Figure 1**, in which participants heard a Midland-accented or Southern-accented talker but were given no dialect information. In this earlier experiment, Southerners shifted in the Southern voice condition with no label just as they did in the present experiment with the Midland label. Non-Southerners exhibited marginal shifts toward the Southern voice that was not labeled in this study, and exhibited no shifts toward the unlabeled Midland voice. These results are consistent with the present study. Non-Southerners clearly shift when a Southern label is given, regardless

of the actual voice they hear, and still shift but only marginally when they must create that label themselves. Southerners shift toward the Southern voice when it is unlabeled but also when it is labeled as a Midland voice. In other words, Non-Southerners tend to attune to the label (even if it is a label they created themselves), while Southerners tend to attune to the voice.

The full set of results from the regression model is shown in **Table 2**. In addition to the two-way interactions observed in **Table 3**, we also see a three-way interaction between Phase, Condition, and Dialect (Est = –0.346, p = 0.022), suggesting that Southerners and Non-Southerners respond differently in their baseline-to-exposure shifts in each condition.

Table 3: Contrasts among estimates, derived from the *emmeans* package.

		Estimate	SE	z-ratio	p-value	
Within Facets	A	–0.210	0.081	–2.583	0.010	**
	B	–0.130	0.083	–1.567	0.117	
	C	0.003	0.089	0.039	0.969	
	D	–0.262	0.089	–2.945	0.003	**
Compare Conditions	A-B (Non-Southerners)	0.08	0.100	0.804	0.421	
	C-D (Southerners)	–0.266	0.111	–2.390	0.017	*
Compare Dialects	A-C (Southern Label)	0.214	0.105	2.045	0.041	*
	B-D (Southern Voice)	–0.132	0.105	–1.251	0.211	

3.1 Self-reported beliefs about talker dialect background

We turn now to participants' self-reported beliefs about the dialect of the model talker they heard and ask whether differences across participant dialect groups may account for the different convergence patterns of Southerners vs. Non-Southerners. **Figure 4** shows participants' self-reported beliefs about where each model talker is from. Participants were asked to rate how likely the voice they heard was to be from the South and the Midwest on a 0–100 point scale, with 100 meaning very likely and 0 meaning not at all likely. We present results for both ratings scales in **Figure 4**, but focus here on the Southernness ratings (left facet), as it is these Southernness ratings that are likely to have an effect on convergence toward monophthongal /aɪ/. Both dialect groups took the Southern voice (left two bars) to be a slightly better indicator that the talker was from the South than the Southern label (right two bars), as shown in the left facet of **Figure 4**, though this difference across conditions does not reach statistical significance (Wilcoxon test: W = 2001.5, p = 0.109). Southernness ratings were generally above the midpoint (see **Table 4**), suggesting that, in either talker condition, social expectations for Southernness were induced. It

is not the case that either group converges toward monophthongal /aɪ/ without also generally reporting that the talker was likely from the South. It is also worth noting that Southerners reported believing the Southern label (mean = 61.2, median = 70) to a comparable extent as Non-Southerners (mean = 66, median = 68).

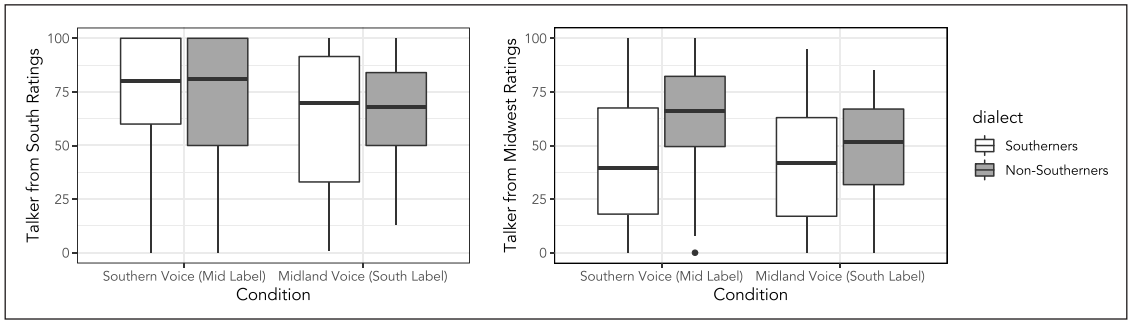

Figure 4: Model talker ratings of likely to be from the South (left) and likely to be from the Midwest (right), across talker conditions and participant dialect background.

Table 4: Talker Southernness Ratings, by participant dialect background and experiment condition.

	Southern Voice	**Southern Label**
Southerners	mean = 73.8, median = 80, sd = 29.5	mean = 61.2, median = 70, sd = 34.7
Non-Southerners	mean = 70.7, median = 81, sd = 31.6	mean = 66, median = 68, sd = 25.8

In general, there was a tendency to rate both talkers as more likely to be Southern (mean likelihood of being from the South rating = 67.9) than Midwestern (mean likelihood of being from the Midwest rating = 49.2). This is likely because the Southern accent is a clear clue to a talker's regional origin, but a more "neutral" Midland-accented talker could also conceivably be from the South, especially if labeled as such. Midwest ratings were generally lower, with the exception being Non-Southerners who were told the Southern-accented talker was from the Midwest, who reported believing the label, perhaps indicating less familiarity with Southern-accented speech. A Wilcoxon test predicting Midwest ratings by dialect background for the Midland Label condition (left two bars of right facet) suggests that the Non-Southerners are marginally more likely to believe the Midland label (W = 291, p-value = 0.05). However, neither "Likely to be from the South" nor "Likely to be from the Midwest" ratings significantly predicted either group of participants' degree of convergence in the Word Naming Game task.

This was determined with a linear regression model predicting exposure-baseline difference, with predictors SouthernnessRating*Dialect and MidwestRating*Dialect. None of these effects were significant for "Likely to be from the South" ratings (Southerners Est. –0.0016, p = 0.405; Non-Southerners: Est. = –0.0014, p = .475) or "Likely to be from the Midwest" ratings (Southerners: Est. –0.0029, p = 0.138; Non-Southerners: –0.001391, p = 0.502), nor any interactions.

3.2 Lexical frequency

Finally, the statistical modeling also reveals predicted effects of duration and frequency. Longer and less frequent tokens have higher /aɪ/ glides, as expected. The model also shows an interaction between Phase and Frequency, suggesting that the shift toward weaker /aɪ/ glides is greater for less frequent words. This result is plotted in **Figure 5**. Note that the effect of frequency on convergence is in the opposite direction of the general effect of frequency on vowel reduction: while in the baseline there is a slight tendency for *more* frequent items to be more reduced, in the exposure condition it is *less* frequent items that are more likely to be reduced (i.e., Southern-like), while high frequency items in the exposure phase are more similar to their high frequency counterparts in the baseline. While a three-way or four-way interaction between Phase, Frequency, and Condition and/or Dialect does not reach significance in the model, we see in **Figure 6** that the Phase*Frequency effect is driven primarily by the two conditions in which significant shifts are observed: Non-Southerners in the Southern Label condition and Southerners in the Southern Voice condition. In other words, the frequency effect is clearest in the two conditions where we observe significant convergence.

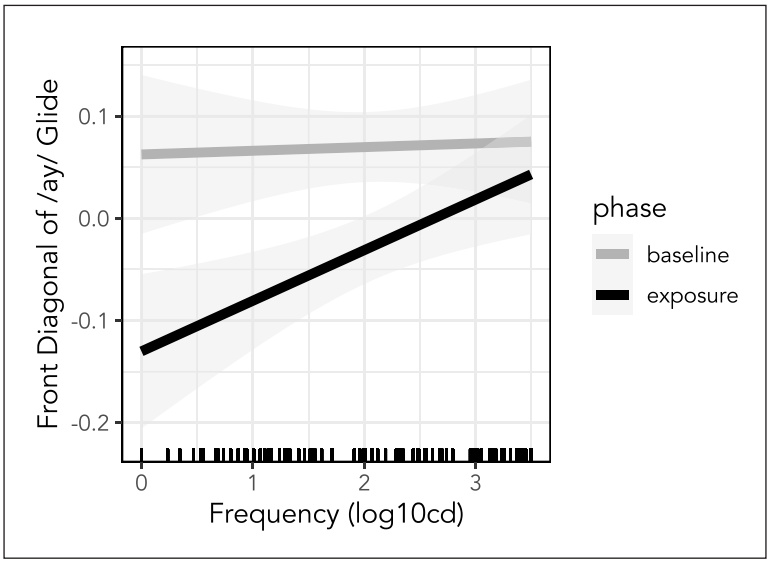

Figure 5: Production of the /aɪ/ glide (at 80%) in the baseline (gray) and exposure (black) phases, as a function of lexical frequency (LOG10CD measure from SUBTLEX-US).

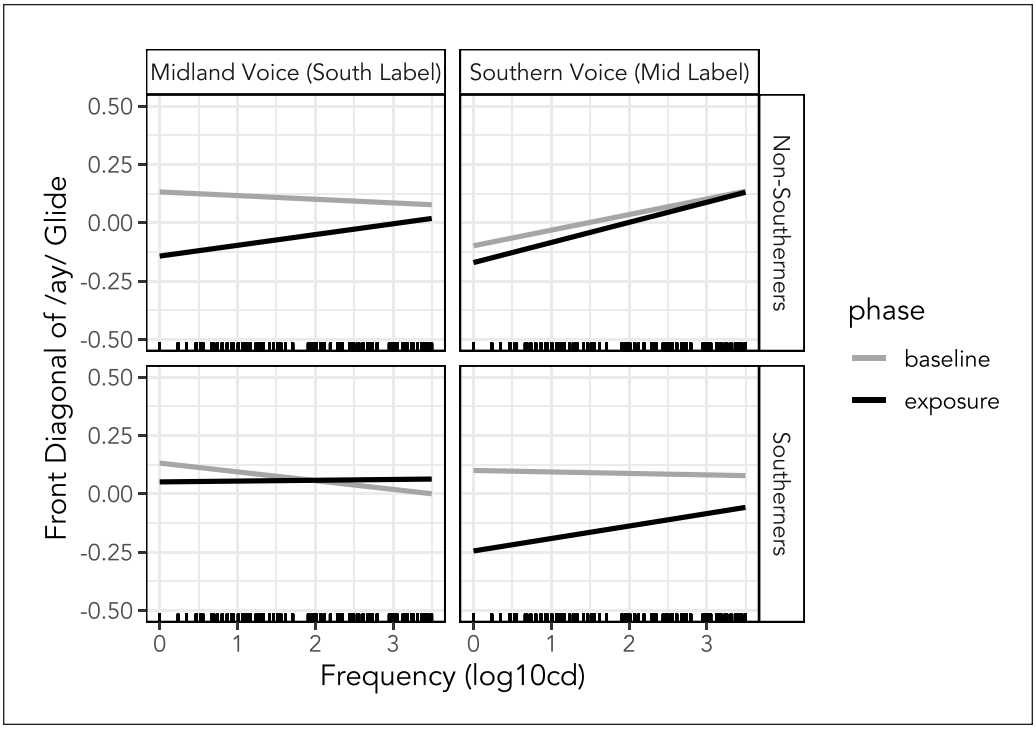

Figure 6: Production of the /aɪ/ glide (at 80%) baseline (gray) and exposure (black) phases, as a function of lexical frequency (lOG10CD measure from SUBTLEX-US), subset by experimental condition and participant dialect background.

4. Discussion

The dialect-label manipulation experiment presented here investigated the associative routes through which variant-to-variant mappings may be activated during expectation-driven convergence. Findings indicate that both direct co-occurrence and indirect socially-mediated routes were utilized in ways that varied systematically by dialect background. When social category labels and acoustic voice information varied independently, Southern participants converged primarily toward the Southern voice while Non-Southerners shifted in response to the Southern label. These results bolster previous findings showing that experience with particular linguistic features not only aids comprehension of those features (Clopper & Pisoni, 2004; McGowan, 2016; Walker, 2018), but also impacts their mental encoding (Sumner & Samuel, 2009) and the extent to which they are adopted into speech production (Love & Walker, 2013; Sanchez et al., 2015; Walker, 2019). In this section, we begin by discussing the different cues Non-Southerners and Southerners use in convergence and suggest that different degrees and types of experience with a given dialect impact the encoding and utilization of sociolinguistic knowledge. We then discuss the implications for models of linguistic cognition more broadly, with a particular focus on the relationship between more implicit convergence shifts which

participants seem to be less aware of and more explicit self-reported beliefs about the model talker's dialect. Finally, we point out the unexpected role of frequency in expectation-driven convergence and discuss some possible explanations for this finding and what it may mean for the mental organization of (socio)linguistic knowledge.

4.1 Dialect differences

Non-Southerners

Differing levels of experience with Southern speech, operationalized here as childhood residential history within or outside of the South, shape not just *degree* of convergence, but also the qualitative patterning of cues to convergence. To start, Non-Southern participants shifted to produce more monophthongal tokens of the /aɪ/ vowel when listening to a talker *labeled* as a Southerner. Since this talker actually had a Midland accent, the social label was the only cue plausibly triggering convergence, suggesting that social labels alone are sufficient in eliciting convergence. Not only does presence of the Southern dialect label facilitate convergence shifts, a Midland label seems to inhibit shifts. These participants did not exhibit similar shifts without the explicit Southern label. When they actually *did* hear a Southern-accented talker who was labeled as being Midland-accented, they failed to shift. The primacy of labels over acoustic cues – at least for Non-Southerners – is in line with previous work showing that social labels can trump linguistic information (Carmichael, 2018; Niedzielski, 1999; Williams, 1989).

How does being provided with an explicit label differ from positing a label based on acoustic cues? When *given* the social category "Southern," Non-Southerners can easily utilize associations between social categories and linguistic forms to generate linguistic expectations. However, when given only acoustic information (as in the case of Wade, 2022), they must first posit the social category "Southern" in order to utilize associative links between Southernness and monophthongal /aɪ/. Even though Non-Southerners *hear* a Southern accent in the Midland label condition, they are also already provided with an explicit label (Midland), so they may be less motivated to posit their own social label from the acoustic input. In other words, if participants are explicitly given social category labels, they have no reason to create these labels on their own. The Midland label therefore seems to disrupt the social-label creation process, blocking the social-category-mediated mapping from observed variants to expected variant. If direct associations between observed Southern variants and monophthongal /aɪ/ do not exist or are weak for Non-Southerners, and if an indirect socially-mediated route is blocked by the "Midland" label, we would expect minimal or no shifts, as we observed in the Midland label condition. We take these findings as evidence that Non-Southerners have socially-mediated mental associations between related Southern variants. We find no evidence that they were utilizing a direct co-occurrence route. If they were, we would expect convergence to the Southern voice regardless of the label (whether created or explicitly given), which we did not find.

Non-Southerners' convergence patterns bolster previous findings that linguistic and social knowledge bidirectionally influence one another: monophthongal /aɪ/ can be predicted from the Southern label, and (as we see in Wade, 2022) the Southern label can be predicted from observing other Southern-accented features. Not only that, but they also suggest that these associative processes may be effectively "stacked," ultimately creating associations (albeit indirect ones) between two variants wherein observing one variant activates social expectations, and these social expectations then activate linguistic expectations for a novel variant. This "stacking" of two bidirectional processes is necessary to account for the behavior of Non-Southerners in Wade (2022), who exhibited marginal shifts toward the same Southern-accented talker who was not assigned a dialect label. If Non-Southerners in fact created their own social category label (e.g., "Southern") after hearing the unlabeled Southern voice in the original experiment, we might expect activation of monophthongal /aɪ/ to be weaker than when they are given the label directly in the present experiment; and indeed, we observe a smaller shift when Non-Southerners must posit their own social category label compared to when they are provided one.

Southerners

On the other hand, Southern participants offer evidence that, among speakers with more sociolinguistic experience, convergence can occur via a more direct variant co-occurrence route, bypassing incongruent social category labels. Southerners were triggered to converge when they heard a Southern-accented talker, suggesting that observed variants were activated and spread activation to non-observed monophthongal /aɪ/ without being impeded by the contradictory "Midland" label. It is our interpretation, then, that a social label need not be activated in these cases, so that linguistic information may be able to trigger expectations for another variant directly. And in fact, a Southern label does not appear to override expectations generated by hearing an incongruent Midland voice, indicating that Southerners generally rely more on direct variant-to-variant associations between Southern features.

The role of experience

We are left with the question of the origin of these differences between dialect groups. Why do Southerners and Non-Southerners utilize different associative routes? We suggest here that the explanation lies in the different types and amounts of experience with Southern speech. Non-Southerners did not exhibit shifts based on acoustic cues signalling Southernness when dialect labels contradicted these cues. We might speculate that this has to do with the relatively greater social salience of monophthongal /aɪ/ compared to other Southern-accented features. Previous research on recognition of the Southern accent has shown that monophthongal /aɪ/ is a particularly salient feature indexing Southernness, while other Southern-accented features may not be as well recognized (Labov, 2010; Plichta & Preston, 2005; Torbert, 2010). Non-Southerners, who

have less experience with Southern speech, may not associate non-/aɪ/ features observed from the Southern-accented talker with Southernness as readily, and therefore may not have formed expectations for monophthongal /aɪ/ from this talker, particularly since some other features of Southern-accented speech, such as back vowel fronting, occur in other dialects of English. The weaker associations between other Southern variants and Southernness may have led to only partial or infrequent activation of this social category based on the Southern voice, which may also explain why Non-Southerners had weaker production shifts than Southerners toward the unlabeled Southern-accented talker in Wade (2022). This weaker activation, combined with a contradictory label, may have been sufficient in more completely blocking associations between these other Southern variants and the social label. However, when provided with the Southern label directly, Non-Southerners were able to bypass this first step mapping other Southern variants to the category "Southern," and were simply left with mapping the category "Southern" to monophthongal /aɪ/, a salient enough association that they should be able to accomplish this even without sufficient experience with the Southern dialect. Indeed, this interpretation aligns with predictions of the dual-route encoding model proposed by Sumner, Kim, King, and McGowan (2014), in which the speech signal is mapped to linguistic and social representations simultaneously. This model suggests that infrequently observed – but strongly socially-weighted – features (like monophthongal /aɪ/) may still be robustly encoded in memory. Southerners, on the other hand, are likely more frequently exposed to co-variation between monophthongal /aɪ/ and other Southern variants, such that it would be easier to establish direct co-occurrence links between monophthongal /aɪ/ and other variants that are not as socially salient.

Our interpretation of these results also supports the out-group homogeneity effect observed in work on group perception and stereotype formation/representation in social psychology. The out-group homogeneity effect essentially finds that out-group individuals are perceived as being more homogeneous than in-group individuals (Park & Rothbart, 1982). Association of monophthongal /aɪ/ with Southernness may also be considered a stereotyped association. Out-group members (here, Non-Southerners) may be expected to adhere more strongly to this stereotype; when presented with a Southern-labeled talker, such stereotyped associations may be activated, and Non-Southerners may expect *any* Southern talker to produce monophthongal /aɪ/. On the other hand, in-group members (here, Southerners) may represent the in-group category "Southern" with more hetereogeneity. For instance, they may recognize that it is only a certain *type* of Southern speaker who may be expected to produce monophthongal /aɪ/, so the "Southern" label would not elicit the expectation for this particular type of talker – only hearing other Southern-accented acoustic features would. It is therefore possible that Southern participants also relied on a socially-mediated category in forming expectations, but that this category was specific to a particular "type" of Southerner who is likely to produce monophthongal /aɪ/, and this association would have to be strong enough so as not to be disrupted by the conflicting Midland label.

Put another way, more experience with Southerners may lead to awareness that simply being from the South does not mean a person will produce monophthongal /aɪ/; rather, a more accurate way to predict monophthongal /aɪ/ is by observing whether *other* features are produced with a Southern accent. Southerners likely have more experience that being from the South does not necessitate Southern-accentedness; they are exposed to frequent counterexamples (and most of these Southern participants were not Southern-shifted themselves). The "Southern" label may not as strongly activate "monophthongal /aɪ/" variants for Southerners because they would presumably have experienced many instances of Southerners *not* using monophthongal /aɪ/, including in their own speech. Further, the concept of Southernness may have more nuanced associations for Southerners, and may therefore not be as strongly linked to this one particular linguistic variant. Non-Southerners, in contrast, may only be aware that a talker is Southern if cued by their accent, and their primary exposure to Southern speech may be (often stereotyped, exaggerated, or parodied) media portrayals of Southerners, which highlight salient features like monophthongal /aɪ/. They would therefore have more associative links between the category "Southerner" and "monophthongal /aɪ/." This interpretation is in line with Drager and Kirtley's (2016) findings that people with military experience don't have stereotyped associations of military speech as sounding Southern, but people without military experience do have these stereotyped associations, due to the common portrayal of members of the U.S. media as having a Southern accent in movies and television. Along these lines, future research could examine whether there is a general tendency for out-group speakers to rely on social category labels, while in-group speakers rely on lower-level acoustic cues in forming expectations about linguistic variants.

4.2 Implicit vs. explicit associations

Although degree of experience plays an important role in the types of cues to convergence speakers use, this does not appear to be a result simply of differences across experience levels in the believability of the dialect labels. After the task, participants rated the talker they heard on how likely they were to have been from the South and the Midwest, on separate 0–100 point scales. "Likelihood of being from the South" ratings were high for both the Southern Voice (Midland Label) and Midland Voice (Southern Label), suggesting that either condition would have been sufficient for eliciting convergence toward Southern speech.

However, it is worth noting that beliefs assessed by these ratings may tap into a different set of expectation-generation mechanisms than expectation-driven convergence shifts. For one, the self-reported assessment of sociolinguistic knowledge we gather from believability ratings differs from the convergence task itself in that it tells us about more *explicit* associations participants may have about a talker and their dialect background. These ratings are presumably susceptible to more controlled, lengthier, introspective reasoning and greater awareness than the quick and

subtle production shifts observed in the Word Naming Game task, which participants did not report being aware of. And in fact, the link between the implicit associations that influence convergence and explicit associations on the ratings task is rather tenuous.

Although Non-Southerners reported marginally higher "likelihood of being from the South" ratings for the Southern voice than for the Southern label, they only converge toward the Southern label – not the Southern voice. Further, self-reported beliefs about the talker's dialect background did not predict degree of convergence for either group of participants. This is not surprising, in light of recent work showing that implicit sociolinguistic associations often do not align with explicitly-reported beliefs. For instance, McGowan and Babel (2020) found that listeners who were told they were listening to either a Spanish or Quechua speaker (but in fact always heard the same speaker), did not shift across guises in an AXB task gauging phonetic category boundaries, yet still reported being influenced by apparent talker dialect in explicit responses after the task. Many others have also found that more implicit and more explicit measures of sociolinguistic behaviors often fail to align (Campbell-Kibler, 2016; D'Onofrio, 2018). We suggest here that Non-Southerners' failure to converge toward the Southern voice, which they rate as more likely to be from the South than the Southern label, is evidence for different processes involved in sociolinguistic cognition, which require differing degrees of awareness, introspection, and conscious control.

4.3 Frequency effects

Finally, we consider how lexical frequency influenced convergence in this study and the implications for theories of speech perception and production. It has been well established in work on imitation that lower-frequency items are more robustly imitated (Babel, 2010; Goldinger, 1998; Goldinger & Azuma, 2004; Nielsen, 2011) (though see Pardo et al. 2013, who failed to replicate these frequency effects). These frequency effects have typically been cited as evidence in support of exemplar models of speech perception and production. The rationale is that lower-frequency items have fewer existing exemplars in memory and are therefore more likely to be influenced by recently heard exemplars; with fewer existing exemplars to compete with, a recently heard token can exhibit a larger effect on the exemplar cloud and therefore a greater influence on the production target. However, we did not expect to observe lexical frequency effects here, since participants never actually hear the target items until they go on to produce them. In other words, our frequency results cannot be straightforwardly explained by reference to a key role for *recently heard episodes of the target items* in a direct perception-production feedback loop. This does not constitute evidence against exemplar-theoretic models per se, since we expect that it is possible to build more complex episodic models to capture expectation-driven convergence (see Goldrick and Cole, 2023, for discussion of possible future directions in this

area). However, these results also invite us to consider alternative explanations for the observed frequency patterns.

We speculate that one possible explanation is that lower-frequency items take longer to access from the production lexicon, leaving more time for integration with expectations about Southern-accentedness. Another related possibility is that lower frequency items are less rote and therefore more available for external influences and generally more variable, due to being less gesturally-practiced. We may also consider functional motivations for convergence and the role of frequency in such motivations. It has been proposed that high frequency items undergo more reduction because they pose less difficulty to the listener in lexical access (e.g., Lindblom, 1990). We would similarly expect low-frequency items to be produced more diphthongally as a form of hyper-speech for items which may pose more difficulty for the listener. However, if we consider matching an interlocutor's speech patterns (or, here, expected speech patterns) to be at least partially functionally motivated by attempts at increasing intelligibility (e.g., Giles et al., 1979; Triandis, 1960), we might actually expect more monophthongization for lower frequency items when conversing with an interlocutor who is expected to be monophthongal. Such motivations may be particularly strong in a game where the end goal is to guess the correct word (though see Kim et al. 2011, for a discussion of how intelligibility-based motivations might lead to *less* convergence). Regardless of the mechanisms behind these frequency effects, they serve to highlight that patterns of behavior around sociolinguistic associations have consequences for more general theories of speech perception and production. Accounting for the role of lexical frequency in convergence patterns, particularly in cases of *expectation-driven* shifts, may be a promising area for future research.

5. Conclusion

We set out to answer the question: what types of mental associations are responsible for the expectation-driven convergence observed by (Wade, 2022), where speakers produce monophthongal /aɪ/ after hearing a Southern talker who does not produce this vowel? By eliciting expectation-driven convergence using a dialect-label manipulation task that isolates the effects of social labels and acoustic cues, we found that the answer varies across participant groups. We suggest that in-group members rely more on direct associations, linking commonly co-occurring Southern variants directly, while out-group members rely on indirect socially-mediated links that require generating a social category label. These results highlight the important role that sociolinguistic experience plays in how sociolinguistic knowledge is formed, represented, and utilized in speech production.

Data accessibility statement

Data and scripts are publicly available at https://osf.io/s8wuh.

Acknowledgments

The authors would like to thank RAs Sadie Butcher and Leila Pearlman for their work on this project, as well as Vanessa Sims for lending her voice. This work was supported by NSF grant BCS-1917900.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

First author: Conceptualization (lead), methodology (lead), data collection, supervision (lead), data analysis and visualization (lead), funding acquisition (lead), writing – original draft (lead), writing – review & editing (lead).

Second author: Conceptualization (supporting), methodology (supporting), writing – original draft (supporting), writing – review & editing (supporting).

Third author: Conceptualization (supporting), methodology (supporting), data analysis (supporting), supervision (supporting), funding acquisition (supporting), writing – original draft (supporting), writing – review & editing (supporting).

References

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(4), 437–456. DOI: https://doi.org/10.1017/S0047404510000400

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1, retrieved 2019 from http://www.praat.org/

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods, 41*(4), 977–90. DOI: https://doi.org/10.3758/BRM.41.4.977

Campbell-Kibler, K. (2012). The implicit association test and sociolinguistic meaning. *Lingua, 122*(7), 753–763. New Horizons in Sociophonetic Variation and Change. DOI: https://doi.org/10.1016/j.lingua.2012.01.002

Campbell-Kibler, K. (2016). Toward a cognitively realistic model of meaningful sociolinguistic variation. In A. Babel (Ed.), *Awareness and control in sociolinguistic research.* Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139680448.008

Carmichael, K. (2018). “Since when does the Midwest have an accent?” The role of regional U.S. accents and reported speaker origin in speaker evaluations. *English World-Wide, 39*(2), 127–156. DOI: https://doi.org/10.1075/eww.00008.car

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61*, 30–47. DOI: https://doi.org/10.1016/j.wocn.2017.01.001

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech, 47*(3), 207–239. DOI: https://doi.org/10.1177/00238309040470030101

Dodsworth, R., & Kohn, M. (2012). Urban rejection of the vernacular: The SVS undone. *Language Variation and Change, 24*(2), 221–245. DOI: https://doi.org/10.1017/S0954394512000105

D’Onofrio, A. (2015). Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics, 19*(2), 241–256. DOI: https://doi.org/10.1111/josl.12115

D’Onofrio, A. (2018). Controlled and automatic perceptions of a sociolinguistic marker. *Language Variation and Change, 30*(2), 261–285. DOI: https://doi.org/10.1017/S095439451800008X

Drager, K., & Kirtley, J. M. (2016). Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In A. Babel (Ed.), *Awareness and control in sociolinguistic research.* Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139680448.003

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Attention, Perception, & Psychophysics, 67*(2), 224–238. DOI: https://doi.org/10.3758/BF03206487

Giles, H., Scherer, K. R., & Taylor, D. M. (1979). Speech markers in social interaction. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 343–381). Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. DOI: https://doi.org/10.1037/0033-295X.105.2.251

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review, 11*(4), 716–722. DOI: https://doi.org/10.3758/BF03196625

Goldrick, M., & Cole, J. (2023). Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics, 99*, 101254. DOI: https://doi.org/10.1016/j.wocn.2023.101254

Hall, J. S. (1942). The phonetics of Great Smoky Mountain speech. *American Speech, 17*(2), 1–110. DOI: https://doi.org/10.2307/487132

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 23*(3), 351–379. DOI: https://doi.org/10.1515/TLR.2006.014

Kim, M., Horton, W. S., & Bradlow, A. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology, 2*, 125–156. DOI: https://doi.org/10.1515/labphon.2011.004

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *Penn Working Papers in Linguistics, 14*(2), Article 12. 23.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*(2), 141–178. DOI: https://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*(2), 262–268. DOI: https://doi.org/10.3758/BF03193841

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Labov, W. (1994). *Principles of linguistic change, vol. 1. Internal factors.* Blackwell.

Labov, W. (2010). *Principles of linguistic change, vol. 3. Cognitive and cultural factors.* Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444327496

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change.* Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110167467

Lai, W. (2021). The online adjustment of speaker-specific phonetic beliefs in multi-speaker speech perception [Doctoral Dissertation]. University of Pennsylvania.

Lenth, R. V. (2021). Emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.0. Retrieved from https://CRAN.R-project.org/package=emmeans

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer. DOI: https://doi.org/10.1007/978-94-009-2037-8

Love, J., & Walker, A. (2013). Football versus football: Effect of topic on /r/ realization in American and English sports fans. *Language and Speech, 56*(4), 443–460. DOI: https://doi.org/10.1177/0023830912453132

McGowan, K. (2015). Social expectation improves speech perception in noise. *Language and Speech, 58* (4), 502–521. DOI: https://doi.org/10.1177/0023830914565191

McGowan, K. (2016). Sounding Chinese and listening Chinese: Awareness and knowledge in the laboratory. In A. Babel (Ed.), *Awareness and control in sociolinguistic research*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139680448.004

McGowan, K. B., & Babel, A. M. (2020). Perceiving isn't believing: Divergence in levels of sociolinguistic awareness. *Language in Society, 49*(2), 231–256. DOI: https://doi.org/10.1017/S0047404519000782

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62–85. DOI: https://doi.org/10.1177/0261927X99018001005

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*(2), 132–142. DOI: https://doi.org/10.1016/j.wocn.2010.12.007

Pardo, J., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*(3), 183–195. DOI: https://doi.org/10.1016/j.jml.2013.06.002

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology, 42*(6), 1051–1068. DOI: https://doi.org/10.1037/0022-3514.42.6.1051

Plichta, B., & Preston, D. R. (2005). The /ay/s have it the perception of /ay/ as a North-South stereotype in United States English. *Acta Linguistica Hafniensia, 37*(1), 107–130. DOI: https://doi.org/10.1080/03740463.2005.10416086

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Reed, P. (2014). Inter- and intra-generational monophthongization and Southern Appalachian identity. *Southern Journal of Linguistics, 38*(1), 159–193.

Sanchez, K., Hay, J., & Nilson, E. (2015). Contextual activation of Australia can affect New Zealanders' vowel productions. *Journal of Phonetics, 48,* 76–95. DOI: https://doi.org/10.1016/j.wocn.2014.10.004

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

Strand, E. (1999). Uncovering the roles of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18*(1), 86–99. DOI: https://doi.org/10.1177/0261927X99018001006

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology, 4*. DOI: https://doi.org/10.3389/fpsyg.2013.01015

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*(4), 487–501. DOI: https://doi.org/10.1016/j.jml.2009.01.001

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America, 128*(4), 2090–9. DOI: https://doi.org/10.1121/1.3467771

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

Triandis, H. C. (1960). Cognitive similarity and communication in dyad. *Human Relations, 13*, 175–183. DOI: https://doi.org/10.1177/001872676001300206

Vaughn, C. & Kendall, T. (2019). Stylistically coherent variants: Cognitive representation of social meaning. *Revista de estudos da linguagem, 27*(4), 1787–1830. DOI: https://doi.org/10.17851/2237-2083.0.0.1787-1830

Wade, L. (2020). The linguistic and the social intertwined: Linguistic convergence toward Southern Speech [Doctoral Dissertation]. University of Pennsylvania.

Wade, L. (2022). Experimental evidence for expectation-driven linguistic convergence. *Language, 98*(1), 63–97. DOI: https://doi.org/10.1353/lan.2021.0086

Walker, A. (2018). The effect of long-term second dialect exposure on sentence transcription in noise. *Journal of Phonetics, 71*, 162–176. DOI: https://doi.org/10.1016/j.wocn.2018.08.001

Walker, A. (2019). The role of dialect experience in topic-based shifts in speech production. *Language Variation and Change, 31*(2), 135–163. DOI: https://doi.org/10.1017/S0954394519000152

Williams, R. (1989). The (mis)identification of regional and national accents of English: Pragmatic, cognitive and social aspects. In O. García & R. Otheguy (Eds.), *A reader in cross-cultural communication* (pp. 55–82). De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110848328.55

Wolfram, W., & Christian, D. (1976). *Appalachian speech*. Center for Applied Linguistics.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics 2008*, 5687–5690. DOI: https://doi.org/10.1121/1.2935783

Zehr, J., & Schwarz, F. (2018). Penncontroller for internet based experiments (ibex). doi:https://doi.org/10.17605/OSF.IO/MD832

Zellou, G., Dahan, D., & Embick, D. (2017). Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition And Neuroscience,* 1–16. DOI: https://doi.org/10.1080/23273798.2016.1275710